

9367

From: Bunner, Bridget
Sent: Monday, May 05, 2003 12:36 PM
To: STIC-Biotech/ChemLib
Subject: sequence search

Hi! I'd like to request a sequence search for case 09/894,912 (no pending search is required):

1. the amino acid sequence of SEQ ID NO: 13

Thanks!

Bridget Bunner

Art Unit 1647
CM1-10D12
(703) 305-7148
mailbox 10B19

CRF

Point of Contact

P. Sheppard

Searcher Telephone number: (703) 308-4499

Phone: _____

Location: _____

Date Picked Up: _____

Date Completed: 5/7/03

Searcher Prep/Review: _____

Clerical: _____

Online time: _____

TYPE OF SEARCH:

NA Sequences: _____

AA Sequences: _____

Structures: _____

Bibliographic: _____

Litigation: _____

Full text: _____

Patent Family: _____

Other: _____

VENDOR/COST (where applic.)

STN: _____

DIALOG: _____

Questel/Orbit: _____

DRLink: _____

Lexis/Nexis: _____

Sequence Sys.: _____

WWW/Internet: _____

Other (specify): _____

GenCore version 5.1.4 p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein --protein search, using sw model

Run on: May 6, 2003, 14:50:18 ; Search time 33 Seconds

(without alignments)
1704.572 Million cell updates/sec

Title: US-09-894-912A-13

Sequence: 1516 1 MHLRLISMLFIINFMVEYI.....QOKKRVODKXKSVSTVH 273

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_todent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriophage:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1505	99.3	272	4	Q9BXK4	Q9BXK4 homo sapien
2	1452	95.8	292	4	Q96K87	Q96K87 homo sapien
3	1153	76.1	217	11	Q9CSB2	Q9CSB2 mus musculu
4	644	42.5	265	11	Q97132	Q97132 mus musculu
5	472	31.1	224	4	Q9UCB2	Q9UCB2 mus sapien
6	192	12.7	1299	5	Q26489	Q26489 spodiopora
7	177	11.7	1101	5	Q264D2	Q264D2 entameoba h
8	176.5	11.6	1376	5	Q852S2	Q852S2 dirosophila
9	176.5	11.6	1679	5	Q24301	Q24301 dirosophila
10	176	11.3	1074	5	Q964D1	Q964D1 entameoba h
11	172	11.3	915	11	Q91VK0	Q91VK0 mus musculu
12	167.5	11.0	926	11	Q35171	Q35171 mus musculu
13	167.5	11.0	932	11	Q62030	Q62030 mus musculu
14	162	10.7	913	4	Q96EP4	Q96EP4 homo sapien
15	161	10.6	440	5	Q18003	Q18003 caenorhabdi
16	159	10.5	503	5	Q9U018	Q9U018 giardia lam

17	158.5	10.5	379	4	Q8WVG4	Q8WVG4 homo sapien
18	157.5	10.4	374	13	Q9W6F8	Q9W6F8 xenopus lae
19	157.5	10.4	803	13	Q42114	Q42114 brachydanio
20	156	10.3	898	5	Q76822	Q76822 brachydanio
21	155	10.2	808	13	Q42113	Q42113 brachydanio
22	153.5	10.1	942	5	Q44762	Q44762 caenorhabdi
23	153.5	10.1	1362	13	Q9PVZ4	Q9PVZ4 xenopus lae
24	151	10.0	288	5	Q76510	Q76510 cryptospori
25	150.5	9.9	827	5	Q44765	Q44765 caenorhabdi
26	149	9.8	1371	11	Q9QWV4	Q9QWV4 rictus sp.
27	148	9.8	548	5	Q9CQ45	Q9CQ45 giardia lam
28	147.5	9.7	802	13	Q9W770	Q9W770 gallus gall
29	146.5	9.7	365	11	Q924Y6	Q924Y6 rictus norv
30	146.5	9.7	807	4	Q9HCB6	Q9HCB6 homo sapien
31	145	9.6	624	4	Q94862	Q94862 homo sapien
32	144.5	9.5	660	5	Q21832	Q21832 cryptospori
33	144.5	9.5	2189	5	Q9H105	Q9H105 elmeria ten
34	144	9.5	807	6	Q9GLX9	Q9GLX9 bos taurus
35	143.5	9.5	220	11	Q93KR2	Q93KR2 mus musculu
36	143.5	9.5	807	11	Q8VCC9	Q8VCC9 mus musculu
37	143	9.4	435	5	Q9CQ41	Q9CQ41 giardia lam
38	142	9.4	378	13	Q9W6F9	Q9W6F9 brachydanio
39	142	9.4	921	5	Q969A3	Q969A3 brachydanio
40	141	9.3	261	5	Q95YF9	Q95YF9 ciona savig
41	140	9.2	259	4	Q8WUR5	Q8WUR5 homo sapien
42	138.5	9.1	1368	5	Q21821	Q21821 caenorhabdi
43	138.5	9.1	1433	5	Q9B1H9	Q9B1H9 anopheles g
44	138.5	9.1	1664	5	Q9TVQ2	Q9TVQ2 caenorhabdi
45	137	9.0	213	11	Q9VLE4	Q9VLE4 mus musculu

ALIGNMENTS

RESULT 1

Q9BXK4 PRELIMINARY: PRT: 272 AA.

AC Q9BXK4 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Thrombospondin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Mao Y., Xie Y., Zhou Z., Zhao W., Zhao S., Wang X., Wang S.,
RA Tang R., Chen X., Wu C.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RC TISSUE=PLACENTA;
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF251057; AA034947.1;
DR EMBL; BC022367; AA022367.1;
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 1.
DR SMART; SM00261; FU. 2.
DR SMART; SM00209; TSP1. 1.
DR PROSITE; PS00092; TSP1. 1.
SQ SEQUENCE 272 AA; 30928 MW; CACAECSB781189 CRC64;
Query Match 99.3%; Score 1505; DB 4; Length 272;
Best Local Similarity 100.0%; Pred. No. 9.9e-129;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 3 HLRRLISMLFIINFMVEYIGSONASRGRRRRHMPNVSGCGGCAATCSVDYGCSCRPRL 62
DB 2 HLRRLISMLFIINFMVEYIGSONASRGRRRRHMPNVSGCGGCAATCSVDYGCSCRPRL 61

```

Oy 63 FFALERIGMKOIGVCLSSCPGSGYGTTRYPDINKCTCKADCDCTCNKFCCKCKSGFYHL 122
Db 62 FFALERIGMKOIGVCLSSCPGSGYGTTRYPDINKCTCKADCDCTCNKFCCKCKSGFYHL 121
Oy 123 LGKCLDNCCEGLANNNHMECVSIHCEVSENNPMSPTCKKGTGFRGRTVRREIIQ 182
Db 122 LGKCLDNCCEGLANNNHMECVSIHCEVSENNPMSPTCKKGTGFRGRTVRREIIQ 181
Oy 183 HPSAKGNLCPPTNETRKTCTVORRKCCKGKRGKRRKCKKKNKESKEAI PDSKSLSS 242
Db 182 HPSAKGNLCPPTNETRKTCTVORRKCCKGKRGKRRKCKKKNKESKEAI PDSKSLSS 241
Oy 243 KEIPEORENKOOQKKRKTKVODKOKSVSYTVH 273
Db 242 KEIPEORENKOOQKKRKTKVODKOKSVSYTVH 272

RESULT 2
Oy 096K87 PRELIMINARY: PRT; 292 AA.
AC 096K87;
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DE 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
DE CDNA FL114440 f15, clone HEMBI000915, weakly similar to
subtilisin-like protease PACE4 precursor (EC 3.4.21.-).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=EMBRIO;
RA Tsogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
Nishikawa T., Nagai K., Sugano S., Takahashi-Fujii A., Hara H.,
Tanase T., Nomura Y., Togiyasu S., Kamai F., Hara R., Takeuchi K.,
Arita M., Nabekura T., Ishii S., Kawai Y., Saito K., Yamamoto J.,
Makamatsu A., Nakamura Y., Nagahara K., Masuo Y., Ohshima A.;
"NEBO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AK027346; BAB55051.1;
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00090; tsp_1; 1.
DR SMART: SM00181; EGF; 1.
DR PROSITE: PS50092; TSP1; 1.
SQ SEQUENCE 292 AA; 33243 MW; 01E2774AC3DA6F8 CRC64;

Query Match 95.8%; Score 1452; DB 4; Length 292;
Best Local Similarity 99.2%; Pred. No. 6, 8e-124;
Matches 262; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

```

```

RESULT 3
Oy 09CSB2 PRELIMINARY: PRT; 217 AA.
AC 09CSB2;
DT 01-JUN-2001 (TREMBlrel. 17, Created)
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
DE 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE 2810459H04RIK protein (Fragment).
GN 2810459H04RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=EMBRIO;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
Fleischmann W., Gaasterland T., Gissi C., King B., Kochiya H.,
Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
Schriml L.M., Staudli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Adono H., Balderelli R., Barsh G.,
Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustincich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilting L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai H., Koltunski S.,
RA Hayashizaki Y.;
"Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690 (2001).
DR EMBL: AK013366; BAB28811.1;
DR MGD: MGI:1920030; 2810459H04RIK.
DR InterPro: IPR002174; Furlin-like.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00090; tsp_1; 1.
DR SMART: SM00261; FRT2.
DR SMART: SM00269; TSP1; 1.
DR PROSITE: PS50092; TSP1; 1.
FT NON_TER 217
FT TER 217
SQ SEQUENCE 217 AA; 24304 MW; 0DCF93BE9FB3FBF7 CRC64;

Query Match 76.1%; Score 1153; DB 11; Length 217;
Best Local Similarity 93.1%; Pred. No. 6, 6e-97;
Matches 201; Conservative 5; Mismatches 10; Indels 0; Gaps 0;

```

```

AC 092132;
DT 01-MAY-1999 (Tremblrel. 10, Created)
DT 01-MAY-1999 (Tremblrel. 10, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE Thrombospondin type 1 domain.
GN R-SPONDIN.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Kamata T., Katsube K., Michikawa M., Yamada M., Mizusawa H.;
RT "R-spondin, a novel thrombospondin type 1 domain gene, expressed in
RT the dorsal neural tube."
RL Submitted (Aug-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB016768; BAA75640.1;
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR000884; TSPL.
DR Pfam: PF00090; esp_1.1.
DR SMART: SM00261; FU; 2.
DR SMART: SM00209; TSPL; 1.
DR PROSITE: PS50092; TSPL; 1.
SQ SEQUENCE 265 AA; 29331 MW; FEEB8964743F5963 CRC64;

Query Match 42.5%; Score 644; DB 11; Length 265;
Best Local Similarity 46.5%; Pred. No. 1,2e-50;
Matches 119; Conservative 36; Mismatches 77; Indels 24; Gaps 6;

QY 6 LISWFIILNFEVYIGSONASRGRRRMRHNPVSGCCGCGCATGSDYNGCLSCPKPLFFA 65
DB 11 VLSWTHIA-----VGRGRI-KGRKQRIISAEGSQACAKGCELCSEVNGCLCKSPKFLIL 63

QY 66 LERIGMKQIGVCLSSCPGYYGTRYPDINKCTCKKAD-CDTFNNNPFCTKSKSFYHLG 124
DB 64 LENDIRQGVCLPSCPGYDANPNPMKCIKCIKEHEAFSINFTCKQELYLHKG 123

QY 125 KCLDNPELEANNHTMECVSYHCEVSEKMPSPCTKKGCTGFKRTETRVREIIGHP 184
DB 124 RCPAPCEBSTAANSTWCEGSPACCESEWSPGCPCKKRLCGFRKSEETRRRVLHAP 183

QY 185 SAKNGLCPPTNETRKYORKKCKGKRGKQ---REKRRKPKGSEKALPDSKSLK 240
DB 184 GGDHTTCSPTETKCTYRRTPEEGQKRRKGQGRNARHAPKNSKE--PRNSNS-- 239

QY 241 SSKETPEOREKNOOK 256
DB 240 -----RRHKGQOQ 247

RESULT 5
QY 09UGB2 PRELIMINARY; PRT; 224 AA.
AC 09UGB2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE DJ824F16.3 (Novel protein similar to mouse thrombospondin type 1
DE domain protein R-spondin) (Fragment).
GN DJ824F16.3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Blakey S.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: AL050325; CAB65783.3;
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR000884; TSPL.
DR SMART: SM00261; FU; 2.
DR SMART: SM00209; TSPL; 1.

```

```

DR PROSITE: PS50092; TSPL; 1.
FT NON_TER 224 224
SQ SEQUENCE 224 AA; 25042 MW; 97D26AD34CDBFE12 CRC64;

Query Match 31.1%; Score 472; DB 4; Length 224;
Best Local Similarity 43.3%; Pred. No. 3,9e-35;
Matches 87; Conservative 36; Mismatches 70; Indels 8; Gaps 5;

QY 29 RRRRRMNPVSGCCGCGCATGSDYNGCLSCPKPLFPALERIGMKQIGVCLSSCPGSGYGT 88
DB 22 RRRKQVGTGLGNCT-GCIICSENGCSTCOORLFLPRREGIRQYCKLHDCPPGYFGI 80

QY 89 RYPDINKCTCKKADCDCTFNKNFCTKCKSGFYHLGKCLDNCEGAEANNHTMECVSYH 148
DB 81 RQGVNRRCKKGATCSGSCSPGFCIRCKRQFYLVKGCILTCPPPTLAHONTRECG--E 138

QY 149 CEVSEKMPSPCTKKGCTGFKRTETRVREIIGHPSAKNGLCPPTNETRKYORKKQ 208
DB 139 CELGFGWGSPTCTHNGKTCGSAWGLESRVREAGHAEATQVLSERKCPIDR-PC- 196

QY 209 KGERG---KGRERKRRKPKNK 226
DB 197 PDERSPQKKGRKDRPRKDR 217

RESULT 6
QY 026489 PRELIMINARY; PRT; 1299 AA.
AC 026489;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, Last annotation update)
DE Endoprotease furin.
GN Furin.
OS Spodoptera frugiperda (Fall armyworm).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Plecoptera; Neoptera; Endopterygota; Lepidoptera; Glossata; Ditrysia;
OC Noctuidae; Noctuidae; Amphipyritinae; Spodoptera.
OX NCBI_TaxID=7108;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=SP9;
RA Cieplik M., Klenk H.;
RT "Cloning and functional characterization of furin from Spodoptera
RT frugiperda (Sf9) cells."
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL: Z68888; CAA93116.1;
DR HSSP: Q99405; IMPT.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR000209; Peptidase_S8.
DR InterPro: IPR002884; P_domain.
DR Pfam: PF01483; P_1.1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISTIN.
DR ProDom: PD000717; P_domain; 1.
DR SMART: SM00261; FU; 10.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
KM PROTEASE.
SQ SEQUENCE 1299 AA; 142020 MW; 4C3799C7B8C572AB CRC64;

Query Match 12.7%; Score 192; DB 5; Length 1299;
Best Local Similarity 27.8%; Pred. No. 5,3e-09;
Matches 63; Conservative 26; Mismatches 78; Indels 60; Gaps 12;

QY 37 NVSGCGCGCATGSD-YNGCLSCPKPLFPALERIGMKQIGVCLSSCPGSGYGTYPDINK 95
DB 795 SVCRPAACATGTSERADGCTGSCNHL-----VLDGTGCMACGSPSHET---EDDM 843

QY 96 CTGKADCTCF--NKNFTCKXSGFYHLGKCLDNCPGGLAEANNHTMECVSYHCEYSE 133
DB 844 CAKHESCDTCCGPGTCCVTCHPSTYALDGRCTVSCPPAYYADKKREK---MRCPVG- 899

```

OY 154 WNPWSPCTKGGTCGFKGTETRVREIIIOHPSAKNLCPP--TNETKRC-TVORRCKCK 209
 DB 900 ---CSTCT-----SAFCLSCPEKMKELKKCKCPVGDSCSA 933
 OY 210 GEGGKGRERKRRKPKN-----GESKE---AIPDSKLSSEKPE 247
 DB 934 GEFVAV---DQCKRCNCPACDSCYGENEGHCLTCPPNLLQDYKCVPE 977

RESULT 7
 ID Q964D2 PRELIMINARY; PRT; 1101 AA.

AC Q964D2; 01-DEC-2001 (TREMBlrel. 19, Created)
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Gal/GalNac lectin 1911.
 GN TGL1.
 OS Entamoeba histolytica.
 OC Eukaryota; Entamoebidae; Entamoeba.
 OX NCBI_TaxID=5759;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=HMI.1MSS;
 RX MEDLINE=21391855; PubMed=11500468;
 RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,
 RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,
 RA Tachibana H.;
 RT "Intermediate subunit of the Gal/GalNac lectin of Entamoeba
 RT histolytica is a Member of a Gene Family Containing Multiple CXXC
 RT Sequence Motifs";
 RL Infect. Immun. 69:5892-5898(2001).
 DR EMBL: AF337950; AAK92161.1;
 DR InterPro: IPR000561; BGF-like.
 DR PROSITE: PS01186; BGF_2; UNKNOWN_1.
 KN Lectin.
 SQ SEQUENCE 1101 AA; 119512 MW; C8B6F5C8DE656AEC CRC64;

Query Match 11.74; Score 177; DB 5; Length 1101;
 Best Local Similarity 26.64; Pred. No. 1e-07;
 Matches 55; Conservative 22; Mismatches 80; Indels 50; Gaps 10;

OY 44 GCGATCSD---YNGCL---SC---KRLFFALF---RIGKQIGVCLSSCGSYGTR 89
 DB 773 GTGOSGSDLSKYPGCKKTDSNDNRGFIYATEGSDGSGSPYSNCTTCTKSNTY-RK 831
 OY 90 YPDINKCTKCAKADCDTCFNKNFCTK-----CKSGFYHLGKCLDNCPEGLEA 136
 DB 832 EBGKNGCAKCDKCATCSDKDTCLTCADPLKVGSKDCKGTYVMSNGEC-----KPC 884
 OY 137 NNHTMECVSIHCEVSEMNPMSPCTKKG-KTC--GFKGTETRVREIIIOHPSAKNLCPP 193
 DB 885 THHCEGSSAAECTVCESDTYKYISNGCNCYVDG-----YDEIKGTCTPC 932
 OY 194 TNETRKTCTVORRCKCKGEGKGRERK 220
 DB 933 TSPCTKCVGKKDCEOEOTGENSEKKK 959

RESULT 8
 ID Q852S2 PRELIMINARY; PRT; 1376 AA.
 AC Q852S2; 01-JUN-2002 (TREMBlrel. 21, Created)
 DT 01-JUN-2002 (TREMBlrel. 21, Last sequence update)
 DE 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE LD30182P.
 GN FUR2.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidae; Drosophilidae; Drosophila.

OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RA Stapleton M., Brokstein P., Hong L., Agbayani A., Carlson J.,
 RA Champe M., Chavez C., Dorsett V., Dresnek D., Fairlan D., Fitse E.,
 RA George R., Gonzalez M., Guarin H., Krommiller B., Li P., Liao G.,
 RA Miranda A., Mungall C.J., Nunoo J., Paclab J., Parages V., Park S.,
 RA Patel S., Phouanavong S., Wan K., Lewis S.E., Rubin G.M.,
 RA Ceiniker S.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AY070553; AAL48024.1;
 SQ SEQUENCE 1376 AA; 149716 MW; B6704BA89A3A888B CRC64;

Query Match 11.64; Score 176.5; DB 5; Length 1376;
 Best Local Similarity 28.04; Pred. No. 1.4e-07;
 Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

OY 3 HURLISWFLILNFMETISQNASRGRMRHNPVNSOGCGCATCSQDY-NGCLSCKPR 61
 DB 734 HLHVID-LAVCLQFCPDGYFENS---RNRTCVP-----CEPNCASCDHPREYTSQDH 783
 OY 62 LFALERIGMKQIGVCLSSCGSYGTRYPDINKCTKCAKADDTGCF--NKVFCCKSGF 119
 DB 784 LWNEHK-----CYSACPUDTET---EDNKCAPCHSTACATCGPTDODCTCRSSR 832
 OY 120 YLHKGCLDNCPEGLEANNHTMECVSIHCEVSEMNPMSPCTKKGTCGFKGTETRV-R 178
 DB 833 YAMONKCLISCPDGFYADKRLKCM-----PCGCGCKTC-----TSNGVCS 873
 OY 179 EIIHPSAKNLCPPNETRKTCTVORRK-CCKGE 211
 DB 874 ECLDNT-----LNKDKCIVSGSGCSESE 899

RESULT 9
 ID Q24301 PRELIMINARY; PRT; 1679 AA.
 AC Q24301; 01-NOV-1996 (TREMBlrel. 01, Created)
 DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE FUR2 protein.
 GN FUR2 OR DFUR2 OR CG4235 OR CG18734.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidae; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Ceiniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazek R.G., Chame M., Pfeiffer B.D.,
 RA Man K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.W., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolintshkov S.,
 RA Borokva D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Deicher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson R., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferrieres S., Fleischmann W.,
 RA Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jaitani M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,

RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kuip D., Lai Z.,
RA Liu X., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liao X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mekulov G., Mlishina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Paclet J.M.,
RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spreading A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodage T., Worley K.C., Xu D., Yang S., Yao Q.A.,
RA Ye J., Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of *Drosophila melanogaster*."; *Science* 287:2185-2195(2000).
RL [2]
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=ISO-1;
RX MEDLINE=92381036; PubMed=1512259;
RA Reebroek A.J., Creemers J.W., Pauli I.G., Kurzik-Dumke U., Rentrop M.,
RA Gaeft E.A., Leunissen J.A., de Ven W.J.;
RT "Cloning and functional expression of Dfurlin2, a subtilisin-like
RT proprotein processing enzyme of *Drosophila melanogaster* with multiple
RT repeats of a cysteine motif."; *J. Biol. Chem.* 267:17208-17215(1992).
RL [3]
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=ISO-1;
RX MEDLINE=95186060; PubMed=7880443;
RA Reebroek A.J., Ayoubi T.A., Creemers J.W., Pauli I.G., Ven W.J.;
RT "The Dfurl2 gene of *Drosophila melanogaster*: genetic organization,
RT expression during embryogenesis, and pro-protein processing activity
RT of its translational product Dfurlin2."; *DNA Cell Biol.* 14:223-234(1995).
RL EMBL: AEO03502; AAF48598.1; -
DR EMBL: L3831; AAF69860.1; -
DR HSP: Q99405; IMP1.
DR MEROPS: S08.049; -
DR FlyBase: FBgn004596; Furl2.
DR InterPro: IPR000561; EGF-like.
DR InterPro: IPR002174; Furlin-like.
DR InterPro: IPR002029; Peptidase_S8.
DR InterPro: IPR002884; P_domain.
DR Pfam: PF01483; P.1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR PRINTS: PR00723; SUBTILISIN.
DR ProDom: PD000717; P_domain; 1.
DR SMART: SM00181; EGF; 1.
DR SMART: SM00261; FU; 10.
DR PROSITE: PS00136; SUBTILASE_ASP; 1.
DR PROSITE: PS00137; SUBTILASE_HIS; 1.
DR PROSITE: PS00138; SUBTILASE_SER; 1.
SQ SEQUENCE 1679 AA; 183368 MW; 3F9E749F0B021CF6 CRC64;

Query Match 11.6%; Score 176.5; DB 5; Length 1679;
Best Local Similarity 28.0%; Pred. No. 1.7e-07;
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

QY 3 HRLISMFLILNFMEXYIGSONASRGROPRMHPVNSGCGCATGATSDY-NGCLSCXKPR 61
DB 1037 HHLVID-LAVVLCQCPDGYFENS---RNRCTVP-----CPNCAQSDHPREVCTSDH 1086
QY 62 LFLAERIGMGIQVCLSCSPGYGTRYPDINKTKRKACDCPF--NKPFTKXSGF 119
DB 1087 LVMEHK-----CYACPLOTYET--EDNKAFCSTCATCATPTDODITRCSR 1135
QY 120 YHLGKCLDNCPEGLNNHTMCEVSIIVCEVSEWNPSPCTKXGKTGKGTETTRV-R 178
DB 1136 YAMONKCLISCPDGFVADKKRLECM-----PCQEGCKTC-----TSGVCS 1176

QY 179 EIIHPSAKGNLCPTNTRKTVORKK-CCKGE 211
DB 1177 ECLQWT-----LWKDKCLVSGSGCSESE 1202

RESULT 10
ID 0964D1 PRELIMINARY; PRT; 1074 AA.
AC 0964D1:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Gal/GalNAc lectin Ig12 (fragment).
GN IGL2.
OS Entamoeba histolytica.
OC Eukaryota; Entamoebidae; Entamoeba.
OX NCBI_TaxID=5759;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HMI:IMSS;
RX MEDLINE=21391855; PubMed=11500466;
RA Cheng X.J., Hughes M.A., Huston C.D., Loftus B., Gilchrist C.A.,
RA Lockhart L.A., Ghosh S., Miller-Sims V., Mann B.J., Petri W.A. Jr.,
RA Tachibana H.;
RT "Intermediate Subunit of the Gal/GalNAc lectin of *Entamoeba*
RT *histolytica* is a Member of a Gene Family Containing Multiple CXXC
RT Sequence Motifs."; *Infect. Immun.* 69:5892-5898(2001).
RL EMBL: AF337951; AAK92362.1; -
DR InterPro: IPR000561; EGF-like.
DR PROSITE: PS01186; EGF_2; UNKNOWN_1.
FT LECTIN.
FT NON_TER 1 1074
FT NON_TER 1074 1074
SQ SEQUENCE 1074 AA; 116782 MW; 1552E2D714EB450F CRC64;

Query Match 11.6%; Score 176; DB 5; Length 1074;
Best Local Similarity 25.6%; Pred. No. 1.3e-07;
Matches 53; Conservative 23; Mismatches 81; Indels 50; Gaps 9;

QY 44 GGCATCSD---YNGC-----LSCKPLFPALF---RIGMKQICVCLSCSPGYGTR 89
DB 757 GTCQSCDLSKYPGCKTDTDTCNVDSRGTGYIATGCSDFSGSPSYNCTGIESNY-PK 815
QY 90 YPDINKTKRKACDCPCFNKFN-----CTKCKSGFYHLGKCLDNCPEGLEA 136
DB 816 EGNKNGCARCDKCATCSDKDTCTDTPLKIGSKCDECKGTGYMSNEC-----KPC 868
QY 137 NNHTMCEVSIIVCEVSEWNPSPCTKXG-KTC--GPKRGTTREVEIIQHPSAKGNLCPP 193
DB 869 TNHCSECSAECTVCESDPTKYVISGNCNACVDGF-----YFDEINGTICPC 916
QY 194 TNERKCTVORKKXCKGGRKKGRERK 220
DB 917 TSPCTKCVGVKDCDEBETGCSERKK 943

RESULT 11
ID 091VK0 PRELIMINARY; PRT; 915 AA.
AC 091VK0:
DT 01-DEC-2001 (TREMblrel. 19, Created)
DT 01-DEC-2001 (TREMblrel. 19, Last sequence update)
DE 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Unknown (protein for MGC:18501).
GN PCSK5.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RP SEQUENCE FROM N.A.
RC TISSUE=BREAST TUMOR;

RA Strausberg R.;
 RL Submitted (Aug-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; BC013068; AAH13068.1; -
 DR MGD; MGI:97515; Pcsk6.
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002029; Peptidase_S8.
 DR InterPro: IPR002884; P_domain.
 DR Pfam; PF01483; P_1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR ProDom; PD000717; P_domain; 1.
 DR SMART; SM00181; EGF; 4.
 DR PROSITE; PS00136; SUBTILASE ASP; UNKNOWN 1.
 DR PROSITE; PS00137; SUBTILASE HIS; UNKNOWN 1.
 DR PROSITE; PS00138; SUBTILASE SER; UNKNOWN 1.
 DR PROSITE; PS00139; SUBTILASE SER; UNKNOWN 1.
 SO SEQUENCE 915 AA; 101488 MW; EEI3FF8FAFC78BE CRC64;

Query Match 11.3%; Score 172; DB 11; Length 915;
 Best Local Similarity 28.4%; Pred. No. 2.5e-07;
 Matches 52; Conservative 26; Mismatches 53; Indels 52; Gaps 13;

QY 36 PNVSO-GCGG-GCATCSDYNGCLSCKRLFFALERIGMKQIGVCLSCPSGYYGTYRPDI 93
 DB 640 PECSEVGCGGPGPHCDSD---CL---HYYYKLK---NNTRICVSSCPGHY---HADK 685
 QY 94 NCTCKKADCDPTCFNK--NCTCKSGFYLN--LKKCLDNGCEGLLEANNHMECVSIHVC 149
 DB 686 KCRKRCAPCESCSHGNOCISCKTGYFUNERTSSCVTQCDDG-----SY 731
 QY 150 EVSEWNPSPCTKKQKTC-GFKGCTETRVREI1QHPSAKGNLCPTNETKCTVORRKCQ 208
 DB 732 EPIKKNVCGKSCENKACIGFHNCTE-----CKGGL---SLQGRCSV---TCE 774
 QY 209 KGE 211
 DB 775 DGO 777

RESULT 12
 ID 035171 PRELIMINARY; PRT; 296 AA.
 AC 035171;
 DT 01-JAN-1998 (Tremblrel. 05, Created)
 DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE PAC4 (Fragment).
 OS PCSK6 OR PAC4.
 GN Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN-BALB/C.
 RA RX MEDLINE=98053888; PubMed=9393739;
 RA Hubbard F.C., Goodrow T.L., Liu S.C., Brilliant M.H., Basset P.,
 RA Waine R.E., Klein-Szanto A.J.;
 RT "Expression of PAC4 in chemically induced carcinomas is associated
 RT with spindle cell tumor conversion and increased invasive ability";
 RL Cancer Res. 57:5226-5231(1997).
 DR EMBL; AF008222; AAB95315.1; -
 DR MEROPS; S08.075; -
 DR MGD; MGI:102897; Pcsk6.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00001; EGF; 1.
 DR SMART; SM00001; EGF; 1.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 DR PROSITE; PS00139; SUBTILASE SER; 1.
 DR PROSITE; PS00139; SUBTILASE SER; 1.
 FT NON TER 1
 SO SEQUENCE 296 AA; 32735 MW; 7796F39216A1E221 CRC64;

Query Match 11.0%; Score 167.5; DB 11; Length 296;

Best Local Similarity 24.1%; Pred. No. 2.1e-07;
 Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

QY 19 YIGSONASRGRRRHRHPRVSGCCGCGATCSDYN--GLSCSKRLFFALERIGMKQIGV 76
 DB 64 YFGDAAARRRR-----CHKGETTGSRPAQCLSCR-RGFY---HHQETNT 106
 QY 77 CLSCPSGYYGTYRPDIINKCTKACADCDPTCFNK-NCTCKSGFYLNHKKCLDNGCEGL 135
 DB 107 CVTLCPAGLYADESQRLL-CLRHPSCKCKVDEPKCTVCKEKFSLIAGSCIPDEPCTY 164
 QY 136 ANNHTEC-----VSIVCEVS---EWNWSPCTKKGCTGFPKGTET 176
 DB 165 FDSLVVCGCHHTCTCGVSRRECHICKAKSFHPDVKCVAPCGE-----GF----- 212
 QY 177 VREI1QHPSAKGNLCPTNETKCTVORRKCQKGE 212
 DB 213 -----YFEMPGLL-PHKVCRCEENCLSCGSSR 240

RESULT 13
 ID 062030 PRELIMINARY; PRT; 932 AA.
 AC 062030;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-MAR-2002 (Tremblrel. 20, Last annotation update)
 DE PAC4A (Fragment).
 GN PCSK6.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=LIVER;
 RA Nakayama K.;
 RL Submitted (MAR-1995) to the EMBL/Genbank/DBJ databases.
 RN (2)
 RP SEQUENCE FROM N.A.
 RC STRAIN=ICR; TISSUE=LIVER;
 RA Hosaka M., Murakami K., Nakayama K.;
 RT "PAC4A is a ubiquitous endoprotease that has similar but not
 RT identical substrate specificity to other Kex2-like processing
 RT endoproteases";
 RL Biomed. Res. 15:383-390(1994).
 DR EMBL; D50060; BAA08777.1; -
 DR HSSP; Q99405; IMPT.
 DR MEROPS; S08.075; -
 DR MGD; MGI:102897; Pcsk6.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR000822; Peptidase_S8.
 DR InterPro; IPR002884; P_domain.
 DR InterPro; IPR000822; ZnF_C2H2.
 DR Pfam; PF01483; P_1.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR PRINTS; PR00723; SUBTILISIN.
 DR PRODOM; PD000717; P_domain; 1.
 DR SMART; SM00181; EGF; 2.
 DR SMART; SM00001; EGF; 1.
 DR SMART; SM00261; FU; 5.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE HIS; 1.
 DR PROSITE; PS00138; SUBTILASE SER; 1.
 DR PROSITE; PS00139; SUBTILASE SER; 1.
 FT NON TER 1
 SO SEQUENCE 932 AA; 103310 MW; 15660179C8CAC72E CRC64;

Query Match 11.0%; Score 167.5; DB 11; Length 932;
 Best Local Similarity 24.1%; Pred. No. 6.5e-07;
 Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

```

OY 19 YIGSONASRGRRORRMHPNVSGCGGATGSDYNG--GCLSCRPFLFALERIGMKOIGV 76
DB 700 YFGAARRRRR-----CKKGETCTGRSPACLSGR-RGFY-----HHQENT 742
OY 77 CLSCSPGVIYGYTRYPINKCTCKKADCDTCFNK-NFTCKCKSGFYHLGKCLDNCPEGLE 135
DB 743 CVTLCPAGLVADESORL--CLRCHPSCKCKVDEPEKCTVKEKESFLARSCIPDEPGTY 800
OY 136 ANNHTEC-----VSIVHCEVS-----EMNWPSPCTKGTGCFKNGTETR 176
DB 801 FDSLVLVCGCHHTCTCTGVPSPRECHICAKSFHQMKVPCVPCGE-----GF----- 848
OY 177 VREIIQHPSAKGNLCPTNETKCTVQKCKQKGER 212
DB 849 -----YPEMPGL--PARYCRCECNCLSCGSSR 876

```

RESULT 14

```

O96EP4
ID O96EP4 * PRELIMINARY; PRT: 913 AA.
AC O96EP4
DT 01-DEC-2001 (TRENBLREL. 19, Created)
DT 01-DEC-2001 (TRENBLREL. 19, Last sequence update)
DT 01-JUN-2002 (TRENBLREL. 21, Last annotation update)
DE Protease convertase subtilisin/kexin type 5.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=KIDNEY;
RA Strausberg R.;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC012064; AAH12064.1;
DR MEROPS: S08.076;
DR InterPro: IPR000561; EGF-like
DR InterPro: IPR000209; Peptidase_S8.
DR InterPro: IPR002884; P_domain.
DR Pfam: PF01483; P_1.
DR Pfam: PF00082; Peptidase_S8; 1.
DR ProDom: PD000717; P_domain; 1.
DR SMART: SMO0181; EGF; 4.
DR PROSITE: PS00136; SUBTILASE_ASP; UNKNOWN_1.
DR PROSITE: PS00137; SUBTILASE_HIS; UNKNOWN_1.
DR PROSITE: PS00138; SUBTILASE_SER; UNKNOWN_1.
DR SEQUENCE 913 AA; 101649 MW; 7A2E63FEC49104BC CRC64;

```

Query Match 10.7%; Score 162; DB 4; Length 913;

Best Local Similarity 22.3%; Pred. No. 2e-06;

Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

```

OY 21 GSONASGRORRMHPNVSGCGGATGSDYNG--CLSCRPFLFALERIGMKOIGV 78
DB 677 GHYHAK-KRCKKAPN-----ESCFSHDDQMSCKGYFL-----NEETNSCV 721
OY 79 SSCPSG-VYGYTRYPINKCTCKKADCDTCFNK-NFTCKCKSGFYHLGKCLDNCPEGLE 137
DB 722 THCPDSDYDPTK--NLCKRCKSENCKTCTEFHNCTECRDGLSOSRCSVSCEDGRYFN 778
OY 138 NHTE-----CVS-----IVHCEVSEW-----NPMSPCTK 163
DB 779 GOCOPCHRFATCAGACADGCTNCTEGYFMEDRCVQSCSISYFPISSSENGYKSKKC 838
OY 164 GKTG-----GFRGTE-----TRREIIQHPSAKGNLCPTNETKCTV 202
DB 839 DISCLTNGPGFKNCTSPSGYLLDLGMQMGAIKDATEESMAEGFCMLVKKNLCL-- 896
OY 203 QKCKCK 209
DB 897 ORKVLQ 903

```

RESULT 15

```

O18003 PRELIMINARY; PRT: 440 AA.
ID O18003
AC O18003
DT 01-JAN-1998 (TRENBLREL. 05, Created)
DT 01-JAN-1998 (TRENBLREL. 05, Last sequence update)
DT 01-DEC-2001 (TRENBLREL. 19, Last annotation update)
DE R17.3 protein.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Pelodierinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN (1)
RP SEQUENCE FROM N.A.
RA Barlow K.;
RL Submitted (MAR-1997) to the EMBL/GenBank/DBJ databases.
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=99063613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
investigating biology."
RT Science 282:2012-2018(1998).
DR EMBL: Z92809; CAB07269.1;
DR InterPro: IPR001212; Somatomedin_B.
DR InterPro: IPR000884; TSP1.
DR Pfam: PF00090; TSP1; 1.
DR Pfam: PF00209; TSP1; 1.
DR SMART: SMO0209; TSP1; 1.
DR PROSITE: PS00524; SOMATOMEDIN_B; 1.
DR PROSITE: PS50092; TSP1; 1.
DR SEQUENCE 440 AA; 50400 MW; 42820B834F263EE6 CRC64;

```

Query Match 10.6%; Score 161; DB 5; Length 440;

Best Local Similarity 23.7%; Pred. No. 1.2e-06;

Matches 53; Conservative 30; Mismatches 87; Indels 54; Gaps 8;

```

OY 59 KRLFLFALERIGMKOIGVCLSCSPSGYGYTRYPINKCTCKKADCDTCFNK-NFTCKCKSG 118
DB 172 KPRHLLIRYSLSKFMPKVTSTPLYEENRVQAPANLYLLESSISECYDEHC----- 225
OY 119 FYHLGKCLDNCPEGLEANNHTECVSIVHCEVSEPMNPMSPCTKCKGKGTETRYR 178
DB 226 --VTIGDC-----CSYTRVCP--RDCVLTDMDSWTOCTADNCTGI--GTOKLR 271
OY 179 EIIQHPSAKGNLCPTNETKCTVQKCKQKGERKCKRERKRPKNGESKEAIPDSKS 238
DB 272 HVIOHARGAACEPLKEMKTCFVE--CR-----PKSALDD-- 306
OY 239 LESSKEIPEORENKQOKKRR-----VODKOKSVSYVH 273
DB 307 ITTVALIDYRNKTRKXIRNNIYWDLPNVAEKKATITYCVH 350

```

Search completed: May 6, 2003, 14:52:39

Job time : 37 secs

GenCore version 5.1.4.P5 4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 14:51:43 ; Search time 22 seconds

(without alignments)
1070.739 Million cell updates/sec

Title: US-09-894-912a-13

Perfect score: 1516
Sequence: 1 MGLRLISWLFITLNFMEYI.....QOKRRYDQKXSVSTVH 273

Scoring table: BLOSUM62
Gapop 10.0, Gapext 0.5

Searched: 328255 seqs, 86286685 residues

Total number of hits satisfying chosen parameters: 328255

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database: Published Applications AA:

1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep.*
2: /cgn2_6/ptodata/1/pubpaa/PT_NEW_PUB.pep.*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep.*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep.*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep.*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep.*
7: /cgn2_6/ptodata/1/pubpaa/PTUS_PUBCOMB.pep.*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep.*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep.*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep.*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep.*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep.*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep.*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1516	100.0	273	US-09-894-912a-13	Sequence 13, Appl
2	1505	99.3	272	US-10-125-852-23	Sequence 23, Appl
3	1505	99.3	272	US-09-894-912a-10	Sequence 10, Appl
4	1505	99.3	272	US-09-894-912a-34	Sequence 34, Appl
5	1472	97.1	265	US-09-894-912a-26	Sequence 26, Appl
6	1472	97.1	292	US-10-125-852-25	Sequence 25, Appl
7	1472	97.1	292	US-09-894-912a-48	Sequence 48, Appl
8	1472	97.1	292	US-09-745-763-166	Sequence 166, Appl
9	1400	92.3	251	US-09-894-912a-16	Sequence 16, Appl
10	1310.5	86.4	279	US-09-894-912a-32	Sequence 32, Appl
11	1275	84.1	225	US-10-185-770-4	Sequence 4, Appl
12	903	49.3	160	US-09-894-912a-14	Sequence 14, Appl
13	656	43.3	263	US-10-125-852-18	Sequence 18, Appl
14	648.5	42.8	243	US-10-125-852-21	Sequence 21, Appl
15	644	42.5	229	US-10-125-852-24	Sequence 24, Appl
16	638	42.1	229	US-09-894-912a-25	Sequence 25, Appl
17	584.5	38.6	243	US-10-185-770-2	Sequence 2, Appl
18	577.5	38.1	243	US-10-125-852-13	Sequence 13, Appl
19	573	37.8	250	US-10-125-852-3	Sequence 3, Appl

20	545	35.9	222	US-10-125-852-15	Sequence 15, Appl
21	540.5	35.7	229	US-10-125-852-6	Sequence 6, Appl
22	276	18.2	46	US-09-894-912a-18	Sequence 18, Appl
23	234.5	15.5	131	US-10-125-852-9	Sequence 9, Appl
24	221	14.6	42	US-09-894-912a-22	Sequence 22, Appl
25	213	14.1	37	US-09-894-912a-20	Sequence 20, Appl
26	202	13.3	110	US-10-125-852-11	Sequence 11, Appl
27	177.5	11.7	43	US-10-125-852-7	Sequence 7, Appl
28	173.5	11.4	969	US-09-961-403-6	Sequence 6, Appl
29	160.5	10.6	479	US-09-764-898-221	Sequence 221, App
30	158.5	10.5	379	US-09-905-291a-4	Sequence 4, Appl
31	158.5	10.5	379	US-10-066-500-22	Sequence 22, Appl
32	158.5	10.5	379	US-09-902-853-4	Sequence 4, Appl
33	158.5	10.5	379	US-09-907-824-4	Sequence 4, Appl
34	158.5	10.5	379	US-09-907-841-4	Sequence 4, Appl
35	158.5	10.5	379	US-09-904-011-4	Sequence 4, Appl
36	158.5	10.5	379	US-10-028-072-308	Sequence 308, App
37	158.5	10.5	379	US-09-906-742-4	Sequence 4, Appl
38	158.5	10.5	379	US-10-121-048-308	Sequence 308, App
39	158.5	10.5	379	US-10-123-904-308	Sequence 308, App
40	158.5	10.5	379	US-10-140-470-308	Sequence 308, App
41	158.5	10.5	379	US-09-906-838-4	Sequence 4, Appl
42	158.5	10.5	379	US-09-907-613-4	Sequence 4, Appl
43	158.5	10.5	379	US-09-907-942-4	Sequence 4, Appl
44	158.5	10.5	379	US-10-175-746-308	Sequence 308, App
45	158.5	10.5	379	US-10-176-918-308	Sequence 308, App

ALIGNMENTS

RESULT 1
US-09-894-912a-13
; Sequence 13, Application US/09894912A
; Publication No. US20030044792A1
; GENE: INFORMATION:
; APPLICANT: Tang et al
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 13
; LENGTH: 273
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912a-13
Query Match 100.0%; Score 1516; DB 9; Length 273;
Best Local Similarity 100.0%; Pred. No. 7.6e-101;
Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MGLRLISWLFITLNFMEYISQNASGRORRMPNVSQCGCGCATCSYNNCLSKCP 60
DB 1 MGLRLISWLFITLNFMEYISQNASGRORRMPNVSQCGCGCATCSYNNCLSKCP 60
QY 1 RLFPALEIRIGMQIVCLSCPSGYGTRPDINKCKKADCTCFNKNFCTCKSGGFY 120
DB 61 RLFPALEIRIGMQIVCLSCPSGYGTRPDINKCKKADCTCFNKNFCTCKSGGFY 120
QY 121 LHLGKCLDNCPEGLNANHMTMECVSIHCEVSEPMNPMSPCTKKKTCGFKGTETRRREI 180

```
Db 121 LHGKLDNCEGLEANNHTECVSIHCEVSENNPMSCTCKGCTGFRGTEVREI 180
Qy 181 IOHPSAKNLCPTNETRKTCTVORRKCCKGGERKKRKKPKGSKSEALPDSKSLE 240
Db 181 IOHPSAKNLCPTNETRKTCTVORRKCCKGGERKKRKKPKGSKSEALPDSKSLE 240
Qy 241 SKEIPEORENKQOQKRRKVDOKKSVSVTVH 273
Db 241 SKEIPEORENKQOQKRRKVDOKKSVSVTVH 273

RESULT 2
US-10-125-852-23
; Sequence 23, Application US/10125852
; Publication No. US20030020341
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1
; FILE REFERENCE: HYS-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; PRIORITY FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO: 23
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-23

Query Match 99.3%; Score 1505; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.6e-100;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HRLISWLFILNFMVEYIGSONASRGRRORRHPNVSOCCGGCATCSDYNGCLSGKPRL 62
Db 2 HRLISWLFILNFMVEYIGSONASRGRRORRHPNVSOCCGGCATCSDYNGCLSGKPRL 61
Qy 63 FFALERIGMKOIGVCLSSCPGSGYGTTRYPDINKCTKACDCTCENKPFCTCKSGFYLIH 122
Db 62 FFALERIGMKOIGVCLSSCPGSGYGTTRYPDINKCTKACDCTCENKPFCTCKSGFYLIH 121
Qy 123 LGKCLDNCEGLEANNHTECVSIHCEVSENNPMSCTCKGCTGFRGTEVREIIO 182
Db 122 LGKCLDNCEGLEANNHTECVSIHCEVSENNPMSCTCKGCTGFRGTEVREIIO 181
Qy 183 HPSAKNLCPTNETRKTCTVORRKCCKGGERKKRKKPKGSKSEALPDSKSLESS 242
Db 182 HPSAKNLCPTNETRKTCTVORRKCCKGGERKKRKKPKGSKSEALPDSKSLESS 241
Qy 243 KEIPEORENKQOQKRRKVDOKKSVSVTVH 273
Db 242 KEIPEORENKQOQKRRKVDOKKSVSVTVH 272

RESULT 3
US-09-894-912a-10
; Sequence 10, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIORITY FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
```

```
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 10
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912a-10

Query Match 99.3%; Score 1505; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.6e-100;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 HRLISWLFILNFMVEYIGSONASRGRRORRHPNVSOCCGGCATCSDYNGCLSGKPRL 62
Db 2 HRLISWLFILNFMVEYIGSONASRGRRORRHPNVSOCCGGCATCSDYNGCLSGKPRL 61
Qy 63 FFALERIGMKOIGVCLSSCPGSGYGTTRYPDINKCTKACDCTCENKPFCTCKSGFYLIH 122
Db 62 FFALERIGMKOIGVCLSSCPGSGYGTTRYPDINKCTKACDCTCENKPFCTCKSGFYLIH 121
Qy 123 LGKCLDNCEGLEANNHTECVSIHCEVSENNPMSCTCKGCTGFRGTEVREIIO 182
Db 122 LGKCLDNCEGLEANNHTECVSIHCEVSENNPMSCTCKGCTGFRGTEVREIIO 181
Qy 183 HPSAKNLCPTNETRKTCTVORRKCCKGGERKKRKKPKGSKSEALPDSKSLESS 242
Db 182 HPSAKNLCPTNETRKTCTVORRKCCKGGERKKRKKPKGSKSEALPDSKSLESS 241
Qy 243 KEIPEORENKQOQKRRKVDOKKSVSVTVH 273
Db 242 KEIPEORENKQOQKRRKVDOKKSVSVTVH 272

RESULT 4
US-09-894-912a-34
; Sequence 34, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIORITY FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 34
; LENGTH: 272
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912a-34

Query Match 99.3%; Score 1505; DB 9; Length 272;
Best Local Similarity 100.0%; Pred. No. 4.6e-100;
Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 3 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATCSQDYNGLCSCKPRL 62
DB 2 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATCSQDYNGLCSCKPRL 61
QY 63 PFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKCTCKACDCTCFNKNFCTCKSGFYHL 122
DB 62 PFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKCTCKACDCTCFNKNFCTCKSGFYHL 121
QY 123 LGKCLDNCEGLANNTMECVSIHCEVSENNPSPCTCKGKGTGFRGTETRVREIIQ 182
DB 122 LGKCLDNCEGLANNTMECVSIHCEVSENNPSPCTCKGKGTGFRGTETRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKTCTVORRKKCKGGERKKRKKPNKSGSKAIPDSKSLSS 242
DB 182 HPSAKGNLCPTNETRKTCTVORRKKCKGGERKKRKKPNKSGSKAIPDSKSLSS 241
QY 243 KEIPEORENKQOQKRRKVODKXSVSTVH 273
DB 242 KEIPEORENKQOQKRRKVODKXSVSTVH 272

RESULT 5
US-09-894-912A-26
; Sequence 26, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 26
; LENGTH: 265
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-894-912A-26

Query Match 97.1%; Score 1472; DB 9; Length 265;
Best Local Similarity 100.0%; Pred. No. 9,9e-98;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATCSQDYNGLCSCKPRL 62
DB 2 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATCSQDYNGLCSCKPRL 61
QY 63 PFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKCTCKACDCTCFNKNFCTCKSGFYHL 122
DB 62 PFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKCTCKACDCTCFNKNFCTCKSGFYHL 121
QY 123 LGKCLDNCEGLANNTMECVSIHCEVSENNPSPCTCKGKGTGFRGTETRVREIIQ 182
DB 122 LGKCLDNCEGLANNTMECVSIHCEVSENNPSPCTCKGKGTGFRGTETRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKTCTVORRKKCKGGERKKRKKPNKSGSKAIPDSKSLSS 242
DB 182 HPSAKGNLCPTNETRKTCTVORRKKCKGGERKKRKKPNKSGSKAIPDSKSLSS 241
QY 243 KEIPEORENKQOQKRRKVODKXSVSTVH 273
DB 242 KEIPEORENKQOQKRRKVODKXSVSTVH 272

DB 242 KEIPEORENKQOQKRRKVODKXSVSTVH 265
RESULT 6
US-10-125-852-25
; Sequence 25, Application US/10125852
; Publication No. US20030032034A1
; GENERAL INFORMATION:
; APPLICANT: Tang, Y. Tom
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR
; FILE REFERENCE: H1S-43A
; CURRENT APPLICATION NUMBER: US/10/125,852
; PRIOR FILING DATE: 2002-08-20
; PRIOR APPLICATION NUMBER: US 60/316,368
; PRIOR FILING DATE: 2001-08-30
; PRIOR APPLICATION NUMBER: US 09/799,451
; PRIOR FILING DATE: 2001-03-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25
; LENGTH: 292
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-125-852-25

Query Match 97.1%; Score 1472; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.1e-97;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 3 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATCSQDYNGLCSCKPRL 62
DB 2 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATCSQDYNGLCSCKPRL 61
QY 63 PFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKCTCKACDCTCFNKNFCTCKSGFYHL 122
DB 62 PFALERIGMKOIGVCLSSCPGSGYGYTRYPDINKCTCKACDCTCFNKNFCTCKSGFYHL 121
QY 123 LGKCLDNCEGLANNTMECVSIHCEVSENNPSPCTCKGKGTGFRGTETRVREIIQ 182
DB 122 LGKCLDNCEGLANNTMECVSIHCEVSENNPSPCTCKGKGTGFRGTETRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKTCTVORRKKCKGGERKKRKKPNKSGSKAIPDSKSLSS 242
DB 182 HPSAKGNLCPTNETRKTCTVORRKKCKGGERKKRKKPNKSGSKAIPDSKSLSS 241
QY 243 KEIPEORENKQOQKRRKVODKXSVSTVH 273
DB 242 KEIPEORENKQOQKRRKVODKXSVSTVH 272

RESULT 7
US-09-894-912A-48
; Sequence 48, Application US/09894912A
; Publication No. US20030044792A1
; GENERAL INFORMATION:
; APPLICANT: Tang et al.
; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
; FILE REFERENCE: 28110/37260A
; CURRENT APPLICATION NUMBER: US/09/894,912A
; PRIOR FILING DATE: 2002-05-10
; PRIOR APPLICATION NUMBER: To be assigned
; PRIOR FILING DATE: 2001-04-05
; PRIOR APPLICATION NUMBER: 60/266,614
; PRIOR FILING DATE: 2001-02-05
; PRIOR APPLICATION NUMBER: 60/215,733
; PRIOR FILING DATE: 2000-06-28
; PRIOR APPLICATION NUMBER: 09/757,562
; PRIOR FILING DATE: 2001-01-09
; PRIOR APPLICATION NUMBER: 09/543,774
; PRIOR FILING DATE: 2000-04-05
; NUMBER OF SEQ ID NOS: 48

SOFTWARE: Patentin version 3.0
SEQ ID NO 48
LENGTH: 292
TYPE: PRT
ORGANISM: Homo sapiens
US-09-894-912A-48

Query Match 97.1%; Score 1472; DB 9; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.1e-97;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRLISWLFILINFEYIGSONASRGRRMRHNPVSOCCGACATCSDYNGCLSCPKRL 62
DB 2 HRLISWLFILINFEYIGSONASRGRRMRHNPVSOCCGACATCSDYNGCLSCPKRL 61
QY 63 FFALEIRGMKQIGVCLSSCPGSGYGTIRYPINCKTKKADCDTCFNKPFCTKCKSGFYHL 122
DB 62 FFALEIRGMKQIGVCLSSCPGSGYGTIRYPINCKTKKADCDTCFNKPFCTKCKSGFYHL 121
QY 123 LGKCLDNCPEGLEANNHMECVSIHCEVSEWNPSPCTKKGKTCGFKRGTEIRVREIIQ 182
DB 122 LGKCLDNCPEGLEANNHMECVSIHCEVSEWNPSPCTKKGKTCGFKRGTEIRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKTCTVQRKKCKGGERKKRKKPKNGESKEALPDSKLESS 242
DB 182 HPSAKGNLCPTNETRKTCTVQRKKCKGGERKKRKKPKNGESKEALPDSKLESS 241
QY 243 KEIPEORENKQOQKKRKYVDKQKS 266
DB 242 KEIPEORENKQOQKKRKYVDKQKS 265

RESULT 8
US-09-745-763-166
Sequence 166, Application US/09745763
Patent No. US20020065394A1
GENERAL INFORMATION:
APPLICANT: Jacobs, Kenneth
McCoy, John M.
Lavallie, Edward R.
Collins-Racie, Lisa A.
Evans, Cheryl
Merberg, David
Trecay, Maurice
Spaulding, Vikki
TITLE OF INVENTION: SECRETED PROTEINS AND POLYNUCLEOTIDES
ENCODING THEM
NUMBER OF SEQUENCES: 219
CORRESPONDENCE ADDRESS:
ADDRESS: Genetics Institute, Inc.
STREET: 87 Cambridgepark Drive
CITY: Cambridge
STATE: MA
COUNTRY: U.S.A.
ZIP: 02140
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/745,763
FILING DATE: 18-Jun-2000
CLASSIFICATION: <Unknown>
ATTORNEY/AGENT INFORMATION:
NAME: Sprunger, Suzanne A.
REGISTRATION NUMBER: 41,333
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617) 498-8284
TELEFAX: (617) 876-5851
INFORMATION FOR SEQ ID NO. 166:
SEQUENCE CHARACTERISTICS:
LENGTH: 292 amino acids

TYPE: amino acid
STRANDEDNESS: <Unknown>
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 166:
US-09-745-763-166

Query Match 97.1%; Score 1472; DB 10; Length 292;
Best Local Similarity 100.0%; Pred. No. 1.1e-97;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRLISWLFILINFEYIGSONASRGRRMRHNPVSOCCGACATCSDYNGCLSCPKRL 62
DB 2 HRLISWLFILINFEYIGSONASRGRRMRHNPVSOCCGACATCSDYNGCLSCPKRL 61
QY 63 FFALEIRGMKQIGVCLSSCPGSGYGTIRYPINCKTKKADCDTCFNKPFCTKCKSGFYHL 122
DB 62 FFALEIRGMKQIGVCLSSCPGSGYGTIRYPINCKTKKADCDTCFNKPFCTKCKSGFYHL 121
QY 123 LGKCLDNCPEGLEANNHMECVSIHCEVSEWNPSPCTKKGKTCGFKRGTEIRVREIIQ 182
DB 122 LGKCLDNCPEGLEANNHMECVSIHCEVSEWNPSPCTKKGKTCGFKRGTEIRVREIIQ 181
QY 183 HPSAKGNLCPTNETRKTCTVQRKKCKGGERKKRKKPKNGESKEALPDSKLESS 242
DB 182 HPSAKGNLCPTNETRKTCTVQRKKCKGGERKKRKKPKNGESKEALPDSKLESS 241
QY 243 KEIPEORENKQOQKKRKYVDKQKS 266
DB 242 KEIPEORENKQOQKKRKYVDKQKS 265

RESULT 9
US-09-894-912A-16
Sequence 16, Application US/09894912A
Publication No. US20030044792A1
GENERAL INFORMATION:
APPLICANT: Tang et al.
TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
FILE REFERENCE: 28110/37260A
CURRENT APPLICATION NUMBER: US/09/894,912A
CURRENT FILING DATE: 2002-05-10
PRIOR APPLICATION NUMBER: To be assigned
PRIOR FILING DATE: 2001-04-05
PRIOR APPLICATION NUMBER: 60/266,614
PRIOR FILING DATE: 2001-02-05
PRIOR APPLICATION NUMBER: 60/215,733
PRIOR FILING DATE: 2000-06-28
PRIOR APPLICATION NUMBER: 09/757,562
PRIOR FILING DATE: 2001-01-09
PRIOR APPLICATION NUMBER: 09/543,774
PRIOR FILING DATE: 2000-04-05
NUMBER OF SEQ ID NOS: 48
SOFTWARE: Patentin version 3.0
SEQ ID NO 16
LENGTH: 251
TYPE: PRT
ORGANISM: Homo sapiens
US-09-894-912A-16

Query Match 92.3%; Score 1400; DB 9; Length 251;
Best Local Similarity 100.0%; Pred. No. 1.3e-92;
Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 QNARSGRRMRHNPVSOCCGACATCSDYNGCLSCPKRLFFALEIRGMKQIGVCLSSCP 82
DB 1 QNARSGRRMRHNPVSOCCGACATCSDYNGCLSCPKRLFFALEIRGMKQIGVCLSSCP 60
QY 83 SGYGTYPINCKTKKADCDTCFNKPFCTKCKSGFYHLGKCLDNCPEGLEANNHME 142
DB 61 SGYGTYPINCKTKKADCDTCFNKPFCTKCKSGFYHLGKCLDNCPEGLEANNHME 120

QY 143 CVAIVCESEWNPSPCTKKGTGFKGTETRVREIIHQHSAGNLCPTNETRKTCTV 202
 DB 121 CVAIVCESEWNPSPCTKKGTGFKGTETRVREIIHQHSAGNLCPTNETRKTCTV 180
 QY 203 ORKCKOGERGKGRERKRRKPKNGSKKAIIPDSKLESSKTIPEORENKOOKKRVOD 262
 DB 181 ORKCKOGERGKGRERKRRKPKNGSKKAIIPDSKLESSKTIPEORENKOOKKRVOD 240
 QY 263 KOKSVSVTVH 273
 DB 241 KOKSVSVTVH 251

RESULT 10
 US-09-894-912A-32
 ; Sequence 32, Application US/09894912A
 ; Publication No. US20030044792A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang et al.
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
 ; FILE REFERENCE: 28110/37260A
 ; CURRENT APPLICATION NUMBER: US/09/894,912A
 ; PRIOR FILING DATE: 2002-05-10
 ; PRIOR APPLICATION NUMBER: To be assigned
 ; PRIOR FILING DATE: 2001-04-05
 ; PRIOR APPLICATION NUMBER: 60/266,614
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: 60/215,733
 ; PRIOR FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: 09/757,562
 ; PRIOR FILING DATE: 2001-01-09
 ; PRIOR APPLICATION NUMBER: 09/543,774
 ; PRIOR FILING DATE: 2000-04-05
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 32
 ; LENGTH: 279
 ; TYPE: PRT
 ; ORGANISM: Mus musculus
 US-09-894-912A-32

Query Match 86.4%; Score 1310.5; DB 9; Length 279;
 Best Local Similarity 87.1%; Pred. No. 3.3e-86;
 Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;
 QY 3 HRLISWLFILNFMETIGSONASRGRRORRMHPNVSOCCGCGCATCSYNGCLSCPKRL 62
 DB 2 HRLISWLFILNFMETIGSONASRGRRORRMHPNVSOCCGCGCATCSYNGCLSCPKRL 61
 QY 63 FFALERIGMKQIGVCLSCSPSGYGTTRYPDINKCTCKADCDTCFNNKFTCKCKSGFYLIH 122
 DB 62 FFALERIGMKQIGVCLSCSPSGYGTTRYPDINKCTCKADCDTCFNNKFTCKCKSGFYLIH 121
 QY 123 LKCLDNCPEGLANNTHECVSIHCEVSEWNPSPCTKKGTGFKGTETRVREIIHQ 182
 DB 122 LKCLDNCPEGLANNTHECVSIHCEVSEWNPSPCTKKGTGFKGTETRVREIIHQ 181
 QY 183 HPSA--KGNLCPTNETRKTCTVORRCKOGERGKGRERKRRKPKNGSKK--AIPDSKS 238
 DB 182 HPSAKGKGNLCPTNETRKTCTVORRCKOGERGKGRERKRRKPKNGSKK--AIPDSKS 241
 QY 239 LESSKEIPEORENK--OQOKKRVKOOK--OKSVSVTVH 273
 DB 242 LESSKEIPEORENK--OQOKKRVKOOK--OKSVSVTVH 279

; TITLE OF INVENTION: ISOLATED HUMAN SECRETED PROTEINS,
 ; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN SECRETED PROTEINS, AND
 ; FILE REFERENCE: CL0001247
 ; CURRENT APPLICATION NUMBER: US/10/185,770
 ; PRIOR FILING DATE: 2002-07-01
 ; PRIOR APPLICATION NUMBER: 60/301,852
 ; PRIOR FILING DATE: 2001-07-02
 ; NUMBER OF SEQ ID NOS: 4
 ; SOFTWARE: FastSeq for Windows Version 4.0
 ; SEQ ID NO 4
 ; LENGTH: 225
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-185-770-4
 Query Match 84.1%; Score 1275; DB 9; Length 225;
 Best Local Similarity 100.0%; Pred. No. 8.9e-84;
 Matches 224; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HRLISWLFILNFMETIGSONASRGRRORRMHPNVSOCCGCGCATCSYNGCLSCPKRL 62
 DB 2 HRLISWLFILNFMETIGSONASRGRRORRMHPNVSOCCGCGCATCSYNGCLSCPKRL 61
 QY 63 FFALERIGMKQIGVCLSCSPSGYGTTRYPDINKCTCKADCDTCFNNKFTCKCKSGFYLIH 122
 DB 62 FFALERIGMKQIGVCLSCSPSGYGTTRYPDINKCTCKADCDTCFNNKFTCKCKSGFYLIH 121
 QY 123 LKCLDNCPEGLANNTHECVSIHCEVSEWNPSPCTKKGTGFKGTETRVREIIHQ 182
 DB 122 LKCLDNCPEGLANNTHECVSIHCEVSEWNPSPCTKKGTGFKGTETRVREIIHQ 181
 QY 183 HPSAKGNLCPTNETRKTCTVORRCKOGERGKGRERKRRKPK 226
 DB 182 HPSAKGNLCPTNETRKTCTVORRCKOGERGKGRERKRRKPK 225

RESULT 12
 US-09-894-912A-14
 ; Sequence 14, Application US/09894912A
 ; Publication No. US20030044792A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Tang et al.
 ; TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH
 ; FILE REFERENCE: 28110/37260A
 ; CURRENT APPLICATION NUMBER: US/09/894,912A
 ; PRIOR FILING DATE: 2002-05-10
 ; PRIOR APPLICATION NUMBER: To be assigned
 ; PRIOR FILING DATE: 2001-04-05
 ; PRIOR APPLICATION NUMBER: 60/266,614
 ; PRIOR FILING DATE: 2001-02-05
 ; PRIOR APPLICATION NUMBER: 60/215,733
 ; PRIOR FILING DATE: 2000-06-28
 ; PRIOR APPLICATION NUMBER: 09/757,562
 ; PRIOR FILING DATE: 2001-01-09
 ; PRIOR APPLICATION NUMBER: 09/543,774
 ; NUMBER OF SEQ ID NOS: 48
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 14
 ; LENGTH: 160
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-09-894-912A-14

Query Match 59.6%; Score 903; DB 9; Length 160;
 Best Local Similarity 100.0%; Pred. No. 1.9e-57;
 Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 96 CTCKADCDTCFNNKFTCKCKSGFYLIHGLKCLDNCPEGLANNTHECVSIHCEVSEWNP 155
 DB 1 CTCKADCDTCFNNKFTCKCKSGFYLIHGLKCLDNCPEGLANNTHECVSIHCEVSEWNP 60

QY 156 PMSFCTKAGTCGFKAGTETRVREIIQHPASAKNLCPTNETRKTCTVORRCKQGERCK 215
 DB 61 PMSFCTKAGTCGFKAGTETRVREIIQHPASAKNLCPTNETRKTCTVORRCKQGERCK 120
 QY 216 GREKRRKPKKESKALIPDSKLSLESSKEIPEORENKQOQ 255
 DB 121 GREKRRKPKKESKALIPDSKLSLESSKEIPEORENKQOQ 160

RESULT 13

US-10-125-852-18
 ; Sequence 18, Application US/10125852
 ; Publication No. US20030032034A1

GENERAL INFORMATION:
 APPLICANT: Tang, Y. Tom
 TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1
 FILE REFERENCE: HYS-43A
 CURRENT APPLICATION NUMBER: US/10/125,852
 CURRENT FILING DATE: 2002-08-20
 PRIOR APPLICATION NUMBER: US 60/316,368
 PRIOR FILING DATE: 2001-08-30
 PRIOR APPLICATION NUMBER: US 09/799,451
 NUMBER OF SEQ ID NOS: 25
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 18
 LENGTH: 263
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-10-125-852-18

Query Match 43.3%; Score 656; DB 9; Length 263;
 Best Local Similarity 45.3%; Pred. No. 1,2e-39;
 Matches 117; Conservative 43; Mismatches 76; Indels 22; Gaps 4;

QY 6 LISMLFIILNFMETIGSONASGRORRHPNVSOGCGCATGSDYNGCLSCPKRLPFA 65
 DB 11 VLSTWTHIA-----SRRGIKRRKORRISAEQSQAACAGCELSGKPKLFI 63
 QY 66 LERIKMOIGVCLSSCPGSGYGTVPDINKCTCKAD-CDTFNKNPCTCKSGFYHLG 124
 DB 64 LERNDRQVGVCLSPGSPGYGTVPDINKCTCKAD-CDTFNKNPCTCKSGFYHLG 123
 QY 125 KCLDNCPEGLEANNHMECVSIHCEVSENNPMSPTCKGKTGFRGTETRVREIIQHP 184
 DB 124 RCPACPEGSSAANGTMECSIPAOCSEMSWPWPCSKKQOLGFRGSEERTRVLHAP 183
 QY 185 SAKGNLCPTNETRKTCTVORRCKQGERCKGERRRKPKKNG-----ESKALIPDSKLS 240
 DB 184 VGDHAAACSDTKETRTCTVRRPCEGOKRRKKGGRRENNANRLARKESKEAGAGR--- 240
 QY 241 SSKETPEORENKQOQKR 258
 DB 241 -----RRKGQOQOQOQ 251

RESULT 14

US-10-125-852-21
 ; Sequence 21, Application US/10125852
 ; Publication No. US20030032034A1

GENERAL INFORMATION:
 APPLICANT: Tang, Y. Tom
 TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR-1
 FILE REFERENCE: HYS-43A
 CURRENT APPLICATION NUMBER: US/10/125,852
 CURRENT FILING DATE: 2002-08-20
 PRIOR APPLICATION NUMBER: US 60/316,368
 PRIOR FILING DATE: 2001-08-30
 PRIOR APPLICATION NUMBER: US 09/799,451
 PRIOR FILING DATE: 2001-03-05

NUMBER OF SEQ ID NOS: 25
 ; SOFTWARE: PatentIn version 3.1
 ; SEQ ID NO 21
 ; LENGTH: 243
 ; TYPE: PRT
 ; ORGANISM: Homo sapiens
 US-10-125-852-21

Query Match 42.8%; Score 648.5; DB 9; Length 243;
 Best Local Similarity 48.1%; Pred. No. 3.7e-39;
 Matches 114; Conservative 39; Mismatches 69; Indels 15; Gaps 3;

QY 27 RGRORRHRPNVSGCGCATGSDYNGCLSCPKRLPFALEIRIGMOIGVCLSSCPGTY 86
 DB 5 KGRORRISAEQSQAACAGCELSGKPKLFIERNDRQVGVCLSPGSPGY 64
 QY 87 GFRYPDINKCTCKAD-CDTFNKNPCTCKSGFYHLGKCLDNCPEGLEANNHMECVS 145
 DB 65 DARNDNKCKICKIEHCEACFSHNFCTCKEGLYLHKRCYPACPEGSSAANGTMECS 124
 QY 146 IHCEVSENNPMSPTCKGKTGFRGTETRVREIIQHPASAKNLCPTNETRKTCTVORR 205
 DB 125 PAOCSEMSWPWPCSKKQOLGFRGSEERTRVLHAPVGDHAAACSDTKETRTCTVRRV 184
 QY 206 KQKGERCKGERRRKPKKNG-----ESKALIPDSKLSLESSKEIPEORENKQOQKR 258
 DB 185 PCPEGOKRRKKGGRRENNANRLARKESKEAGAGR-----RRKGQOQOQOQ 231

RESULT 15

US-10-125-852-24
 ; Sequence 24, Application US/10125852
 ; Publication No. US20030032034A1

GENERAL INFORMATION:
 APPLICANT: Tang, Y. Tom
 TITLE OF INVENTION: METHODS AND MATERIALS RELATING TO NOVEL STEM CELL GROWTH FACTOR
 FILE REFERENCE: HYS-43A
 CURRENT APPLICATION NUMBER: US/10/125,852
 CURRENT FILING DATE: 2002-08-20
 PRIOR APPLICATION NUMBER: US 60/316,368
 PRIOR FILING DATE: 2001-08-30
 PRIOR APPLICATION NUMBER: US 09/799,451
 NUMBER OF SEQ ID NOS: 25
 SOFTWARE: PatentIn version 3.1
 SEQ ID NO 24
 LENGTH: 265
 TYPE: PRT
 ORGANISM: Mus musculus
 US-10-125-852-24

Query Match 42.5%; Score 644; DB 9; Length 265;
 Best Local Similarity 46.5%; Pred. No. 8.4e-39;
 Matches 119; Conservative 36; Mismatches 77; Indels 24; Gaps 6;

QY 6 LISMLFIILNFMETIGSONASGRORRHPNVSOGCGCATGSDYNGCLSCPKRLPFA 65
 DB 11 VLSTWTHIA-----VGRGI-KGRORRISAEQSQAACAGCELSGKPKLFI 63
 QY 66 LERIKMOIGVCLSSCPGSGYGTVPDINKCTCKAD-CDTFNKNPCTCKSGFYHLG 124
 DB 64 LERNDRQVGVCLSPGSPGYGTVPDINKCTCKAD-CDTFNKNPCTCKSGFYHLG 123
 QY 125 KCLDNCPEGLEANNHMECVSIHCEVSENNPMSPTCKGKTGFRGTETRVREIIQHP 184
 DB 124 RCPACPEGSSAANGTMECSIPAOCSEMSWPWPCSKKQOLGFRGSEERTRVLHAP 183
 QY 185 SAKGNLCPTNETRKTCTVORRCKQGERCKGERRRKPKKNG-----ESKALIPDSKLS 240
 DB 184 GGDHTTSDTKETRTCTVRRPCEGOKRRKKGGRRENNANRLARKESKEAGAGR--- 239
 QY 241 SSKETPEORENKQOQKR 256

Db 240 -----RRHKG000 247
|:|:|:
Search completed: May 6, 2003, 14:53:51
Job time : 23 secs

GenCore version 5.1.4 p5 4578
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 14:50:58 ; Search time 15 seconds
(without alignments)

535,497 Million cell updates/sec

Title: US-09-894-912A-13

Perfect score: 1516
Sequence: 1 MGLRLISWLFILNFMET.....QQRKRVQDKXSVSTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database: Issued Patents AA:*

1: /cgn2_6/prodata/1/aa/5A.COMB.pep:.*
2: /cgn2_6/prodata/1/aa/5B.COMB.pep:.*
3: /cgn2_6/prodata/1/aa/6A.COMB.pep:.*
4: /cgn2_6/prodata/1/aa/6B.COMB.pep:.*
5: /cgn2_6/prodata/1/aa/6C.COMB.pep:.*
6: /cgn2_6/prodata/1/aa/backfile1.pep:.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed.
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	173.5	11.4	969	2	US-08-284-941-2
2	173.5	11.4	969	2	US-08-447-642-2
3	173.5	11.4	969	4	US-09-236-503-2
4	173.5	11.4	969	5	PCT-US93-02147A-2
5	162	10.7	799	2	US-08-525-940-23
6	162	10.7	799	2	US-08-976-838-23
7	162	10.7	881	2	US-08-525-940-21
8	162	10.7	881	2	US-08-976-838-21
9	162	10.7	915	2	US-08-525-940-18
10	162	10.7	915	2	US-08-976-838-18
11	162	10.7	915	4	US-09-214-5558-2
12	162	10.7	915	4	US-09-214-5558-7
13	161	10.6	288	1	US-08-368-852-15
14	157.5	10.4	288	2	US-08-525-940-15
15	157.5	10.4	288	1	US-08-976-838-15
16	149	9.8	568	2	US-07-862-021B-14
17	149	9.8	568	5	PCT-US93-03164-14
18	148.5	9.8	2523	1	US-08-185-432-18
19	148.5	9.8	2523	4	US-08-899-232-18
20	147.5	9.7	802	1	US-07-862-021B-12
21	147.5	9.7	802	1	US-08-313-288B-12
22	147.5	9.7	802	5	PCT-US93-03164-12
23	146	9.6	1068	1	US-08-537-210A-2
24	146	9.6	1068	4	US-09-113-825-2
25	146	9.6	2556	4	US-08-185-432-17
26	146	9.6	2556	4	US-08-899-232-2
27	143.5	9.5	807	1	US-07-862-021B-10

28	143.5	9.5	807	1	US-08-313-288B-10	Sequence 10, Appl
29	143.5	9.5	807	5	PCT-US93-03164-10	Sequence 10, Appl
30	139	9.2	366	4	US-08-857-076-103	Sequence 103, App
31	139	9.2	486	3	US-08-746-559A-5	Sequence 5, Appl1
32	139	9.2	516	3	US-08-746-559A-4	Sequence 4, Appl1
33	139	9.2	1367	2	US-08-249-687C-2	Sequence 2, Appl1
34	139	9.2	1367	2	US-08-625-819-2	Sequence 2, Appl1
35	139	9.2	1367	3	US-08-746-559A-2	Sequence 2, Appl1
36	139	9.2	1367	4	US-08-864-641B-18	Sequence 18, Appl
37	139	9.2	2556	4	US-08-083-590A-20	Sequence 20, Appl
38	139	9.2	2556	3	US-08-532-384-20	Sequence 20, Appl
39	135.5	8.9	370	4	US-08-857-076-104	Sequence 104, App
40	135.5	8.9	1382	2	US-08-737-715-2	Sequence 2, Appl1
41	135.5	8.9	1382	4	US-09-457-040B-7	Sequence 7, Appl1
42	130	8.6	2471	1	US-08-185-432-16	Sequence 16, Appl1
43	130	8.6	2471	1	US-08-083-590A-19	Sequence 19, Appl
44	130	8.6	2471	3	US-08-532-384-19	Sequence 19, Appl
45	130	8.6	2471	4	US-08-899-232-1	Sequence 1, Appl1

ALIGNMENTS

RESULT 1
US-08-284-941-2
; Sequence 2, Application US/08284941
; Patent No. 5863756
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PAGE 4 AND
TITLE OF INVENTION: PAGE 4.1 GENE AND POLYPEPTIDES IN CELLS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSER: COOLEY GODWARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/284,941
FILING DATE: 2 August 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: NEBLEY PH.D., RICHARD L.
REGISTRATION NUMBER: 30092
REFERENCE/DOCKET NUMBER: CHIR-009/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 843-5070
TELEFAX: (415) 857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 969 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-284-941-2
Query Match 11.4%; Score 173.5; DB 2; Length 969;
Best Local Similarity 28.8%; Pred. No. 8e-07;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;
QY 19 YIGQVNASGRGRORRMPHNVSOQCGCATGCD--YNGCLCKPRLFFALERIGMGIV 76
Db 737 YFEDTAAKRCR-----CHKGCTGSSRAATQCLSCR-RGFY-----HHQEMNT 779

QY 77 CLSCPSGYYGTRYPDINKTKCKADCDTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
DB 780 CVTLCPAGFYADE--SQKNCKLKCHPSCKKCVDEPEKCTVCKEGFSLARGSCIPDCEPGTY 837
QY 136 ANNHTECVSIVHCEVSEMNPMSPCTKKGKTCG 168
DB 838 FDSLELRGCECHH-----TCG 853

RESULT 2
US-08-447-642-2
Sequence 2, Application US/08447642
Patent No. 5989890
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J
APPLICANT: KIEFER, MICHAEL C
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PAGE 4 AND
TITLE OF INVENTION: PAGE 4.1 GENE AND POLYPEPTIDES IN CELLS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODDARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/447,642
FILING DATE: 23-MAY-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/284,941
FILING DATE: 2 August 1994
ATTORNEY/AGENT INFORMATION:
NAME: NEELEY PH.D., RICHARD L.
REGISTRATION NUMBER: 30092
REFERENCE/DOCKET NUMBER: CHIR-009/01US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 843-5070
TELEFAX: (415) 857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 969 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-447-642-2

Query Match 11.4%; Score 173.5; DB 2; Length 969;
Best Local Similarity 28.8%; Pred. No. 8e-07;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 19 YIGSONASRGRRORHMHNVNSQCGGCATCSD--YNGCLSKPRLPFLALRIGMKOIGV 76
DB 737 YFGDTARCR-----CHKCETCSSRAATGCLSR-RGFV-----HHQEWNT 779
QY 77 CLSCPSGYYGTRYPDINKTKCKADCDTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
DB 780 CVTLCPAGFYADE--SQKNCKLKCHPSCKKCVDEPEKCTVCKEGFSLARGSCIPDCEPGTY 837
QY 136 ANNHTECVSIVHCEVSEMNPMSPCTKKGKTCG 168
DB 838 FDSLELRGCECHH-----TCG 853

RESULT 3
US-09-236-503-2

Sequence 2, Application US/09236503
Patent No. 6277590
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J
APPLICANT: KIEFER, MICHAEL C
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PAGE 4 AND 4.1 GENE AND
TITLE OF INVENTION: POLYPEPTIDES IN CELLS
FILE REFERENCE: CHIR-009/0405
CURRENT APPLICATION NUMBER: US/09/236,503
EARLIER FILING DATE: 1999-01-25
EARLIER APPLICATION NUMBER: 08/447,642
EARLIER FILING DATE: 1995-05-23
EARLIER APPLICATION NUMBER: 08/284,941
EARLIER FILING DATE: 1994-08-02
EARLIER APPLICATION NUMBER: 07/848,629
NUMBER OF SEQ ID NOS: 16
SOFTWARE: Patent In Ver. 2.0
SEQ ID NO 2
LENGTH: 969
TYPE: PRT
ORGANISM: Homo sapiens
US-09-236-503-2

Query Match 11.4%; Score 173.5; DB 4; Length 969;
Best Local Similarity 28.8%; Pred. No. 8e-07;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

QY 19 YIGSONASRGRRORHMHNVNSQCGGCATCSD--YNGCLSKPRLPFLALRIGMKOIGV 76
DB 737 YFGDTARCR-----CHKCETCSSRAATGCLSR-RGFV-----HHQEWNT 779
QY 77 CLSCPSGYYGTRYPDINKTKCKADCDTCFNK-NFCTKCKSGFYHLGKCLDNCPEGLE 135
DB 780 CVTLCPAGFYADE--SQKNCKLKCHPSCKKCVDEPEKCTVCKEGFSLARGSCIPDCEPGTY 837
QY 136 ANNHTECVSIVHCEVSEMNPMSPCTKKGKTCG 168
DB 838 FDSLELRGCECHH-----TCG 853

RESULT 4
PCT-US93-02147A-2
Sequence 2, Application PC/TUS9302147A
GENERAL INFORMATION:
APPLICANT: BARR, PHILIP J
APPLICANT: KIEFER, MICHAEL C
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR PAGE 4 AND
TITLE OF INVENTION: PAGE 4.1 GENE AND POLYPEPTIDES IN CELLS
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: COOLEY GODDARD CASTRO HUDDLESON & TATUM
STREET: FIVE PALO ALTO SQUARE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94306
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/02147A
FILING DATE: 19930309
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/848,629
FILING DATE: 09-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: NEELEY PH.D., RICHARD L.
REGISTRATION NUMBER: 30092
REFERENCE/DOCKET NUMBER: CHIR-009/00US

TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 494-7622
TELEFAX: (415) 857-0663
TELEX: 380816 COOLEY PA
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 969 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US93-02147A-2

Query Match 11.4%; Score 173.5; DB 5; Length 969;
Best Local Similarity 28.8%; Pred. No. 8e-07;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

Db 19 YIGSNASRGRRMRHPNVSGGCGCATGSPD--YNGCLSCDPLFPALERIGMKGICV 76
737 YGDDMARCR-----CHKQCTCSSRAATQCLSCR-RGFR-----HDEBMT 779

Qy 77 CLSCPSGYGTRYPDKCTCKADCDTCFNNK-NFTCKKSGFYHLGKCLDNCEGLE 135
780 CVTLPPAGFYADE--SOKNCLKCHPSCKKCVDEPEKCTVCEGEGSLARGSCIPDCEPGTY 837

Qy 136 ANNHTMECVSIHCEVSENNPMSPTCKKGKTCG 168
838 FDESLIRGCECH-----TCG 853

Db 838 FDESLIRGCECH-----TCG 853

RESULT 5
US-08-525-940-23
Sequence 23, Application US/08525940
Patent No. 5866351
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
APPLICANT: Wolf, Joseph R.
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,940
FILING DATE:

CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/368, 852
FILING DATE: 01-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,322
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 799 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 6.1e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

Qy 21 GSONASRGRRMRHPNVSGGCGCATGSPDYN--CLSCDPLFPALERIGMKGICV 78
Db 563 GHYADK-KKCRKCAPN-----CESCFSGHODQCKCKGYFL-----NETNSCV 607

Qy 79 SCSPSG-YGTRYPDKCTCKADCDTCFNNK-NFTCKKSGFYHLGKCLDNCEGLE 137
Db 608 THCPDGSYDPTK--NLCKKSEKCTCTEFNCTECRDGLSLQSCRSVCEDEGRYFN 664

Qy 138 NHTME-----CVS-----IYHCEVSENNPMSPTCKK 163
Db 665 GDDQCPHRCATCAGAGDCTINCTEGYFMEDRCVOSISYTFHDSSENGYCKKC 724

Qy 164 GKTG-----GPKKGT-----TRVEIIQHPAKGNLCPTNETRKTCTV 202
Db 725 DISLCTNGGPFNCTSCPSGYLLDLGCMGAIKDATEESWAEQFCMLVKKNNLC-- 782

Qy 203 GRKCK 209
Db 783 QKRYLQ 789

RESULT 6
US-08-976-838-23
Sequence 23, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 799 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-23

Query Match 10.7%; Score 162; DB 2; Length 799;
Best Local Similarity 22.3%; Pred. No. 6.1e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

Oy	21	GSQNSRSRRQBRPMHNPVSGQGGCATGSDVNG--CLSCPRFLFALERIKMGIGVCL	78
Db	563	GHVHADR-KACRCRCAVN-----CESCGSHGQDQMSCKRYGYFL-----NETNISCV	607
Oy	79	SSCPSSG-YVGGRRYPDINKTKCKADCPITCFNNKPFCKCKSGFYVHLGKCLDNCPEGLEAN	137
Db	608	THCPGOSVQDTPKR--NLCKKSENMCKTKEPFINCTEGRDGLSLQSGRCSVSGEDGRYFN	664
Oy	138	NHTME-----CVS-----IYVCEVSEW-----NPMSPCTCKK	164
Db	665	GDDCQPCPRFCATCAGAGDGCINTETBEYFMEDGRCVGSCISVYPHSESNQYKSCXKC	724
Oy	164	GRTC-----GRKRGK-----TRVEIIIDHSAKGLCPPTNERRKCTV	202
Db	725	DISLCTLCNGPFRKCTKSCFSGVLLDLMGCMGAIKCDATIEBSMAEGGFCMLVKNKLLC--	782
Oy	203	QRKKCKQK 209	
Db	783	QRKVLDQ 789	

```

1 RESULT 7
2 US-08-525-940-21
3 Sequence 21, Application US/08525940
4 Patent No. 5866351
5
6 GENERAL INFORMATION:
7 APPLICANT: Franzusoff, Alex
8 APPLICANT: Miranda, Luis R.
9 APPLICANT: Wolf, Joseph R.
10 TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
11 TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES
12 NUMBER OF SEQUENCES: 25
13 CORRESPONDENCE ADDRESS:
14 ADDRESSEE: Sheridan Ross & McIntosh
15 STREET: 1700 Lincoln Street, Suite 3500
16 CITY: Denver
17 STATE: Colorado
18 COUNTRY: U.S.A.
19 ZIP: 80203
20
21 COMPUTER READABLE FORM:
22 MEDIUM TYPE: Floppy disk
23 COMPUTER: IBM PC compatible
24 OPERATING SYSTEM: PC-DOS/MS-DOS
25 SOFTWARE: Patent In Release #1.0, Version #1.25
26
27 CURRENT APPLICATION DATA:
28 APPLICATION NUMBER: US/08/525,940
29
30 FILING DATE:
31 CLASSIFICATION: 514
32 PRIOR APPLICATION DATA:
33 APPLICATION NUMBER: US 08/368,852
34 FILING DATE: 01-JAN-1995
35 PRIOR APPLICATION DATA:
36 APPLICATION NUMBER: US 08/088,322
37 FILING DATE: 07-JUL-1993
38 ATTORNEY/AGENT INFORMATION:
39 NAME: Connell, Gary J
40 REGISTRATION NUMBER: 32,020
41 REFERENCE/DOCKET NUMBER: 2848-11-C1
42 TELECOMMUNICATION INFORMATION:
43 TELEPHONE: (303) 863-9700
44 TELEFAX: (303) 863-0223
45 INFORMATION FOR SEQ ID NO: 21:
46 SEQUENCE CHARACTERISTICS:
47 LENGTH: 881 amino acids
48 TYPE: amino acid
49 TOPOLOGY: linear
50 MOLECULE TYPE: protein
51
52 US-08-525-940-21

```

Query Match	10.7%;	Score 162;	DB 2;	Length 881;
Best Local Similarity	22.3%;	Pred. No. 6.9e-06;		
Matches	55;	Conservative	33;	Mismatches 81;
				Indels 78;
				Gaps 12.

[illegible]

```

US-08-976-838-21
: Sequence 21, Application US/08976838
: Patent No. 5981259
:
: GENERAL INFORMATION:
: APPLICANT: Franzsoff, Alex
: TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
: TITLE OF INVENTION: MOLECULES
: NUMBER OF SEQUENCES: 31
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sheridan Rosa P.C.
: STREET: 1700 Lincoln St., Suite 3500
: CITY: Denver
: STATE: Colorado
: COUNTRY: U.S.A.
: ZIP: 80203
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/976,838
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Connell, Gary J.
: REGISTRATION NUMBER: 32,020
: REFERENCE/DOCKET NUMBER: 2848-11-C2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (303) 863-9700
: TELEFAX: (303) 863-0223
: INFORMATION FOR SEQ ID NO: 21:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 881 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
:
US-08-976-838-21

```

	Query Match	Score 162;	DB 2;	Length 881;
	Best Local Similarity	10.7%;	Pred. No. 6.9e-06;	
	Matches	55;	Conservative 33;	Mismatches 81; Indels 78; Gaps 12.
Oy	21	GASMAKRRQRYVNHNRVSSGSGCAGTSCSPYUN--CLSCRPFLALERIGMKOIGVCL	78	
		:::::	:::::	
Db	645	GAYHNAIK-RKSKAKRN-----CESCFSHDDQSMCKKYUFL-----NEETNSCV	689	
Oy	79	SACPSG-YUGRTYDILKTKCAADRDIPNKMNCTSKSGFYINGLKDLNCEVGEAN	137	
		:::::	:::::	
Db	690	TTRPDSTVDITRK---NLKRKSENNKITFTFNHTCTCRGLSLDQSRKSYCSDEGRYFN	746	

QY 138 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 163
DB 747 GDDCPCHRFCAATCAGAGADGCTGCTGTFMEDRCVOSCSISYFPHSSENGYKSCKK 806
QY 164 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPTNETRKTCTV 202
DB 807 DISCLTNGPFPKNTSCPSGYLLDLGMCWGAICXDATESWAEQGFCLVKKNLNC-- 864
QY 203 ORKKCOK 209
DB 865 QKKVLQ 871

RESULT 9
US-08-525-940-18
Sequence 18, Application US/08525940
Patent No. 5866351
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,940
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/368,852
FILING DATE: 01-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,322
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-18

Query Match 10.7%; Score 162; DB 2; Length 915;
Best Local Similarity 22.3%; Pred. No. 7, 1e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
QY 21 GSONASRGRORRMHPNVSGCGGATCSDYNG--CLSCRPRLFALERIGMKQIGVCL 78
DB 679 GHYHADK-KRCKKCAPN-----CESCFGSHGQCMCKYGYFL-----NEETNSCV 723
QY 79 SSCPSG-YVGTVPDINKTKCKADCDTCFNNKFPCTCKCKSGFYLLGKCLDNCPEGLEAN 137
DB 724 THCPDGSYODTKK--NLCKKCSGNCKTCTEPHNTCTEGRDGLSLQSGRCSVSCEDGRYFN 780

QY 138 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 163
DB 761 GDDCPCHRFCAATCAGAGADGCTGCTGTFMEDRCVOSCSISYFPHSSENGYKSCKK 840
QY 164 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPTNETRKTCTV 202
DB 841 DISCLTNGPFPKNTSCPSGYLLDLGMCWGAICXDATESWAEQGFCLVKKNLNC-- 898
QY 203 ORKKCOK 209
DB 899 QKKVLQ 905

RESULT 10
US-08-976-838-18
Sequence 18, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 18:
SEQUENCE CHARACTERISTICS:
LENGTH: 915 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-18

Query Match 10.7%; Score 162; DB 2; Length 915;
Best Local Similarity 22.3%; Pred. No. 7, 1e-06;
Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;
QY 21 GSONASRGRORRMHPNVSGCGGATCSDYNG--CLSCRPRLFALERIGMKQIGVCL 78
DB 679 GHYHADK-KRCKKCAPN-----CESCFGSHGQCMCKYGYFL-----NEETNSCV 723
QY 79 SSCPSG-YVGTVPDINKTKCKADCDTCFNNKFPCTCKCKSGFYLLGKCLDNCPEGLEAN 137
DB 724 THCPDGSYODTKK--NLCKKCSGNCKTCTEPHNTCTEGRDGLSLQSGRCSVSCEDGRYFN 780
QY 138 NHTME-----CVS-----IVHCEVSEW-----NPMSPCTKK 163
DB 781 GDDCPCHRFCAATCAGAGADGCTGCTGTFMEDRCVOSCSISYFPHSSENGYKSCKK 840
QY 164 GKTG-----GFKRGTE-----TRVREIIQHPSAKGNLCPTNETRKTCTV 202
DB 841 DISCLTNGPFPKNTSCPSGYLLDLGMCWGAICXDATESWAEQGFCLVKKNLNC-- 898

```

1 RESULT 12
2 US-09-214-555B-7
3 Sequence 7, Application US/09214555B
4 Patent No. 6380171
5 GENERAL INFORMATION:
6 APPLICANT: INSTITUT DE RECHERCHE CLINIQUE DE MONTREAL
7 TITLE OF INVENTION: PRO-PROTEIN CONVERTING ENZYME
8 FILE REFERENCE: PRO-PROTEIN CONVERT ENZ
9 CURRENT APPLICATION NUMBER: US/09/214,555B
10 CURRENT FILING DATE: 1999-01-04
11 PRIOR APPLICATION NUMBER: 60/021,008
12 PRIOR FILING DATE: 1996-07-26
13 PRIOR APPLICATION NUMBER: 2 203,745
14 PRIOR FILING DATE: 1997-04-25
15 NUMBER OF SEQ ID NOS: 9
16 SOFTWARE: PatentIn Ver. 2.1
17 SEQ ID NO: 915
18 LENGTH: 915
19 TYPE: PRT
20 ORGANISM: Homo sapiens
21 US-09-214-555B-7

```

```
Query Match      10.6%; Score 161; DB 1; Length 288;
Best Local Similarity 28.2%; Pred.No.2.4e-06;
Matches 46; Conservative .52; Indels 40; Gaps 11.

QY    36 PNVSQ-GCGG-CGATCSDNYNGLCKSPFLPALERISMGIQVCLSSCPGGYGRYPDI 93
      | :|:|:| | | |:| || ::|:|:: :|||:| | :|
```

Db 13 PECSEVCGDGPDPHCHND---CL---HYYYKLK---NNTRICVSSCPGHHY---HADK 58
QY 94 NKCTKCKADCDTCFNK--NFCTKCKSGFYH--LGKCLDNCEGIEANNHTMECVSIYHC 149
Db 59 KRCRCAPNCSCSCFSGSHDQCMSCSKYGFPLNETNSCVTHCPDGSYDPTKKNLC----- 112
QY 150 EVSEWNPMSPTCKKGTCTGFRGTETRV-REIIQHPASAGNLCPTNETRKTCTVORRKKCO 191
Db 113 -----RKCSXNKTCFTEFHKTCTEER-----DGLSLGGSRC 142

RESULT 14
US-08-525-940-15
Sequence 15, Application US/08525940
Patent No. 5866351
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
APPLICANT: Wolf, Joseph R.
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASES AND GENES
TITLE OF INVENTION: ENCODING SAID PROTEASES
NUMBER OF SEQUENCES: 25
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/525,940
FILING DATE:
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/368,852
FILING DATE: 01-JAN-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/088,322
FILING DATE: 07-JUL-1993
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-525-940-15

Query Match 10.4%; Score 157.5; DB 2; Length 288;
Best Local Similarity 26.6%; Pred. No. 4.7e-06;
Matches 49; Conservative 30; Mismatches 64; Indels 41; Gaps 12;

QY 36 PNVSO-GCGG-GCATCDYNGCLSCPKRLFPALERIGMKQIGVCLSCPSGSGYGTTPDI 93
Db 13 PECSEVCGDGPDPHCHND---CL---HYYYKLK---NNTRICVSSCPGHHY---HADK 58
QY 94 NKCTKCKADCDTCFNK--NFCTKCKSGFYH--LGKCLDNCEGIEANNHTMECVSIYHC 149
Db 59 KRCRCAPNCSCSCFSGSHDQCMSCSKYGFPLNETNSCVTHCPDGSYDPTKKNLC----- 112
QY 150 EVSEWNPMSPTCKKGTCTGFRGTETRV-REIIQHPASAGNLCPTNETRKTCTVORRKKCO 208
Db 113 -----RKCSXNKTCFTEFHKTCTEER-----DGLSLGGSRC 142

Db 113 -----RKCSXNKTCFTEFHKTCTEER-----DGLSLGGSRC 142
QY 209 KGER 212
Db 159 PCHR 162

RESULT 15
US-08-976-838-15
Sequence 15, Application US/08976838
Patent No. 5981259
GENERAL INFORMATION:
APPLICANT: Franzusoff, Alex
APPLICANT: Miranda, Luis R.
APPLICANT: Wolf, Joseph R.
TITLE OF INVENTION: CD4+ T-LYMPHOCYTE PROTEASE NUCLEIC ACID
TITLE OF INVENTION: MOLECULES
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross P.C.
STREET: 1700 Lincoln St., Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,838
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2848-11-C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 288 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-976-838-15

Query Match 10.4%; Score 157.5; DB 2; Length 288;
Best Local Similarity 26.6%; Pred. No. 4.7e-06;
Matches 49; Conservative 30; Mismatches 64; Indels 41; Gaps 12;

QY 36 PNVSO-GCGG-GCATCDYNGCLSCPKRLFPALERIGMKQIGVCLSCPSGSGYGTTPDI 93
Db 13 PECSEVCGDGPDPHCHND---CL---HYYYKLK---NNTRICVSSCPGHHY---HADK 58
QY 94 NKCTKCKADCDTCFNK--NFCTKCKSGFYH--LGKCLDNCEGIEANNHTMECVSIYHC 149
Db 59 KRCRCAPNCSCSCFSGSHDQCMSCSKYGFPLNETNSCVTHCPDGSYDPTKKNLC----- 112
QY 150 EVSEWNPMSPTCKKGTCTGFRGTETRV-REIIQHPASAGNLCPTNETRKTCTVORRKKCO 208
Db 113 -----RKCSXNKTCFTEFHKTCTEER-----DGLSLGGSRC 142

QY 209 KGER 212
Db 159 PCHR 162

Search completed: May 6, 2003, 14:53:24
Job time: 17 secs

```

OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodinae; Xenopus.
CX NCBI_TaxID=8355;
RN RP
RP SEQUENCE FROM N.A.
RC TISSUE=embryo;
RX MEDLINE=93376785; PubMed=8367492;
RA Ruiz J, Altaba A., Cox C., Jessell T.M., Klar A.;
RT "Ecopic neural expression of a floor plate marker in frog embryos
RT injected with the midline transcription factor Pinalavis.";
RL Proc. Natl. Acad. Sci. U.S.A. 90:8268-8272(1993).
CC - FUNCTION: PROMOTES THE ATTACHMENT OF SPINAL CORD AND SENSORY
CC NEURON CELLS AND THE OUTGROWTH OF NEURITES IN VITRO. MAY
CC CONTRIBUTE TO THE GROWTH AND GUIDANCE OF AXONS IN BOTH THE SPINAL
CC CORD AND THE PNS.
CC - SUBCELLULAR LOCATION: Secreted.
CC - TISSUE SPECIFICITY: EXPRESSED AT HIGH LEVELS IN THE FLOOR PLATE.
CC - SIMILARITY: CONTAINS 6 TSP TYPE-1 DOMAINS.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, L09123; AAA19105.1; -.
DR PIR, A47723; A47723.
DR InterPro; IPR002861; Reeler.
DR InterPro; IPR000884; TSP1.
DR Pfam; PF00090; tsp.1; 6.
DR Pfam; PF002014; Reeler.1.
DR SMART; SM00209; TSP1.6.
DR PROSITE; PS50092; TSP1.6.
KW Glycoprotein; Signal; Repeat; Cell adhesion.
KW SIGNAL
KW CHAIN
FT FT 24 803
FT FT 437 488
FT FT 496 548
FT FT 553 604
FT FT 609 661
FT FT 665 716
FT FT 751 803
FT FT 210 210
FT FT 677 677
SQ SEQUENCE 803 AA; 90702 MW; D3A3E329548AED9 CMC64;
Query Match 10.3%; Score 156; DB 1; Length 803;
Best Local Similarity 24.6%; Pred. No. 0.00048;
Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;
OY 38 VSQGGGCGCATCTSPYNGCLSCRPRLFLFALTEIKGKQIGVCLSSCP--GYVGTYPYDINK 95
DB 544 VNEEEPPSSCIYVEMWMEGCS-----ATCMGMKKRHMIMIKGTPADQSMCKADYVEVK 598
OY 96 C--TKCK-----ADDDTCFNQNFCTCKCKSGFLHGLKGLDKNCPEGLANN--- 138
DB 599 CMPEPCHTITPCVLSPSEWSDSVTCGGRTRROR--MLKSPSLSGCNLEDELVKQEV 655
OY 139 -HTMECVSIYHCEVSEWNPWSPCTKKQKTCGFKRGRTETVREIIQHPSAKNIICPTNET 197
DB 656 CMLPSCP--ISCELTETMSYWSGC--NNSCG--KGMIMITRMITMEPFGCAVCEPTVOR 708
OY 198 RKCTVORKCKCKRGKRGKREKKRKKPKQKSKKALIPDKSLSSKELPEQREKQ 253
DB 709 KCKRC--RKCK-----SSGNERHLK-----DKREKRREKIKESDQEO 747

```

ID NTC2 MOUSE STANDARD; PRT: 2470 AA.
 AC 035516; 060941; 006008;
 DT 15-JUN-2002 (Rel. 41, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Neurogenic locus notch homolog protein 2 precursor (Notch 2) (Notch
 B)
 GN NOTCH2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NX NCBI_TaxID=10090;
 RN
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Thymus;
 RA Hamada Y., Higuchi M., Tsujimoto Y.;
 RT "Complete amino acid sequence and multiform transcripts encoded by a
 RT single copy of mouse Notch2 gene."
 RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
 RN
 RP SEQUENCE OF 316-1518 FROM N.A.
 RC STRAIN=C57BL/6 X CBA; TISSUE=Embryo;
 RA Lardelli M., Lendahl U.;
 RT "Notch A and Notch B-two mouse Notch homologues coexpressed in a
 RT wide variety of tissues."
 RL Exp. Cell Res. 204:364-372(1993).
 RN
 RP [3]
 RP SEQUENCE OF 1765-2153 FROM N.A.
 RC MEDLINE=97075110; PubMed=8917536;
 RA Milner L.A., Bigas A., Kopan R., Brahmeh-Stein C., Bernstein I.D.,
 RA Martin D.I.;
 RT "Inhibition of granulocytic differentiation by mNotch1."
 RL Proc. Natl. Acad. Sci. U.S.A. 93:13014-13019(1996).
 RN
 RP [4]
 RP FUNCTION.
 RA Hamada Y., Kadokawa Y., Okabe M., Ikawa M., Coleman J.R.,
 RA Tsujimoto Y.;
 RT "Mutation in ankyrin repeats of the mouse Notch2 gene induces early
 RT embryonic lethality."
 RL Development 126:3415-3424(1999).
 RN
 RP [5]
 RP DEVELOPMENTAL STAGE, AND ALTERNATIVE SPLICING.
 RA Higuchi M., Kiyama H., Hayakawa T., Hamada Y., Tsujimoto Y.;
 RT "Differential expression of Notch1 and Notch2 in developing and adult
 RT mouse brain."
 RL Brain Res. Mol. Brain Res. 29:263-272(1995).
 RN
 RP [6]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RA Saxena M.T., Schroeter E.H., Mumm J.S., Kopan R.;
 RT "Murine notch homologs (N1-4) undergo presenilin-dependent
 RT proteolysis."
 RL J. Biol. Chem. 276:40268-40273(2001).
 RN
 RP [7]
 RP POST-TRANSLATIONAL PROCESSING, AND MUTAGENESIS OF MET-1699.
 RA MEDLINE=21373376; PubMed=11559941;
 RT Mizutani T., Taniguchi Y., Aoki T., Hashimoto N., Honjo T.;
 RT "Conservation of the biochemical mechanisms of signal transduction
 RT among mammalian Notch family members."
 RL Proc. Natl. Acad. Sci. U.S.A. 98:9026-9031(2001).
 CC
 CC -1- FUNCTION: Functions as a receptor for membrane-bound ligands
 CC Jagged2 and Delta to regulate cell-face determination.
 CC Upon ligand activation through the released notch intracellular
 CC domain (NICD) it forms a transcriptional activator complex with
 CC RBP-J kappa and activates genes of the enhancer of split locus.
 CC Affects the implementation of differentiation, proliferation and
 CC apoptotic programs (by similarity). May play an essential role in
 CC postimplantation development, probably in some aspect of cell
 CC specification and/or differentiation.
 CC
 CC -1- SUBUNIT: Heterodimer of a C-terminal fragment N(TM) and a N-

CC terminal fragment N(EC) which are probably linked by disulfide
 CC bonds.
 CC
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein. Following
 CC proteolytical processing NICD is translocated to the nucleus.
 CC
 CC -1- ALTERNATIVE PRODUCTS: 2 isoforms; 1 (shown here) and 2, may be
 CC produced by alternative splicing.
 CC
 CC -1- TISSUE SPECIFICITY: Expressed in the brain, liver, kidney,
 CC neuroepithelia, somites, optic vesicles and branchial arches, but
 CC not heart.
 CC
 CC -1- DEVELOPMENTAL STAGE: Expressed in the embryonic ventricular zone,
 CC the postnatal ependymal cells, and the choroid plexus throughout
 CC embryonic and postnatal development.
 CC
 CC -1- PTM: Synthesized in the endoplasmic reticulum as an inactive form
 CC which is proteolytically cleaved by a furin-like convertase in the
 CC trans-Golgi network before it reaches the plasma membrane to yield
 CC an active, ligand-accessible form. Cleavage results in a C-
 CC terminal fragment N(TM) and a N-terminal fragment N(EC). Following
 CC ligand binding, it is cleaved by TNF-alpha converting enzyme
 CC (TACE) to yield a membrane-associated intermediate fragment called
 CC notch extracellular truncation (NEXT). This fragment is then
 CC cleaved by presenilin dependent gamma-secretase to release a
 CC notch-derived peptide containing the intracellular domain (NICD)
 CC from the membrane.
 CC
 CC -1- PTM: Phosphorylated.
 CC
 CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
 CC
 CC -1- SIMILARITY: CONTAINS 34.5 EGF-LIKE DOMAINS.
 CC
 CC -1- SIMILARITY: CONTAINS 2 LIN/NOTCH REPEATS.
 CC
 CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
 CC
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL, restriction
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC -----
 CC EMBL: D32210; BAA22094.1; -
 CC EMBL: X68279; CAA48340.1; -
 CC EMBL: U1881; AAC52924.1; -
 CC HSSP: P16109, 1PSB.
 CC
 CC WGD: MG1:97364, Notch2.
 CC InterPro: IPR002110; ANK.
 CC InterPro: IPR000152; Asx_hydroxyl.
 CC InterPro: IPR000561; EGF-like.
 CC InterPro: IPR000742; EGF_2.
 CC InterPro: IPR01881; EGF_Ca.
 CC InterPro: IPR001438; EGF_11.
 CC InterPro: IPR008800; Notch.
 CC Pfam: PF00008; EGF_35.
 CC Pfam: PF00023; ank; 6.
 CC Pfam: PF00066; notch; 2.
 CC PRINTS: PR0010; EGFBL00D.
 CC PRINTS: PR01452; NOTCH.
 CC SMART: SM00248; ANK; 4.
 CC SMART: SM00179; EGF_Ca; 22.
 CC SMART: SM00001; EGF-like; 12.
 CC SMART: SM00004; NL_3.
 CC PROSITE: PSS0085; ANK_REPEAT; 4.
 CC PROSITE: PSS0297; ANK_REPEAT; 4.
 CC PROSITE: PSS0010; ASX_HYDROXYL; 22.
 CC PROSITE: PSS0022; EGF_1; 33.
 CC PROSITE: PSS0186; EGF_2; 27.
 CC PROSITE: PSS0187; EGF_Ca; 22.
 CC Receptor: Transcription regulation; Activator; Differentiation;
 CC Transmembrane protein; Repeat; ANK repeat; EGF-like domain;
 CC Transmembrane; Glycoprotein; Signal; Phosphorylation;
 CC Alternative splicing
 CC
 CC SIGNAL 1 25
 CC CHAIN 26 2470
 CC CHAIN 1656 2470
 CC CHAIN 1657 2470
 CC DOMAIN 26 1677
 CC
 CC POTENTIAL.
 CC NEUROGENIC LOCUS NOTCH HOMOLOG PROTEIN 2.
 CC NOTCH EXTRACELLULAR TRUNCATION.
 CC NOTCH INTRACELLULAR DOMAIN.
 CC EXTRACELLULAR (POTENTIAL).


```

FT TRANSMEM 1678 1698 POTENTIAL.
FT DOMAIN 1699 2470 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 26 63 EGF-LIKE 1.
FT DOMAIN 64 102 EGF-LIKE 2.
FT DOMAIN 105 143 EGF-LIKE 3.
FT DOMAIN 144 180 EGF-LIKE 4.
FT DOMAIN 182 219 EGF-LIKE 5.
FT DOMAIN 221 256 EGF-LIKE 6 (INCOMPLETE).
FT DOMAIN 258 294 EGF-LIKE 7.
FT DOMAIN 296 334 EGF-LIKE 8.
FT DOMAIN 336 372 EGF-LIKE 9.
FT DOMAIN 373 411 EGF-LIKE 10.
FT DOMAIN 413 452 EGF-LIKE 11.
FT DOMAIN 454 490 EGF-LIKE 12.
FT DOMAIN 492 528 EGF-LIKE 13.
FT DOMAIN 530 566 EGF-LIKE 14.
FT DOMAIN 568 603 EGF-LIKE 15.
FT DOMAIN 605 641 EGF-LIKE 16.
FT DOMAIN 643 678 EGF-LIKE 17.
FT DOMAIN 680 716 EGF-LIKE 18.
FT DOMAIN 718 753 EGF-LIKE 19.
FT DOMAIN 755 791 EGF-LIKE 20.
FT DOMAIN 793 829 EGF-LIKE 21.
FT DOMAIN 831 869 EGF-LIKE 22.
FT DOMAIN 871 907 EGF-LIKE 23.
FT DOMAIN 909 945 EGF-LIKE 24.
FT DOMAIN 947 983 EGF-LIKE 25.
FT DOMAIN 985 1021 EGF-LIKE 26.
FT DOMAIN 1023 1059 EGF-LIKE 27.
FT DOMAIN 1061 1097 EGF-LIKE 28.
FT DOMAIN 1099 1145 EGF-LIKE 29.
FT DOMAIN 1147 1183 EGF-LIKE 30.
FT DOMAIN 1185 1221 EGF-LIKE 31.
FT DOMAIN 1223 1260 EGF-LIKE 32.
FT DOMAIN 1262 1300 EGF-LIKE 33.
FT DOMAIN 1302 1345 EGF-LIKE 34.
FT DOMAIN 1372 1410 EGF-LIKE 35.
FT DOMAIN 1418 1454 LIN/NOTCH 1.
FT REPEAT 1501 1533 LIN/NOTCH 2.
FT REPEAT 1535 1569 ANK 1.
FT REPEAT 1571 1605 ANK 2.
FT REPEAT 1607 1641 ANK 3.
FT REPEAT 1643 1677 ANK 4.
FT REPEAT 1679 1713 ANK 5.
FT REPEAT 1715 1749 ANK 6.
FT DOMAIN 1751 1785 POLY-ALA.
FT DOMAIN 1787 1821 POLY-LIU.
FT DOMAIN 1823 1857 POLY-ALA.
FT DOMAIN 1859 1893 POLY-SER.
FT DOMAIN 1895 1929 POLY-GLY.
FT DISULFID 1931 1965 BY SIMILARITY.
FT DISULFID 1967 2001 BY SIMILARITY.
FT DISULFID 2003 2037 BY SIMILARITY.
FT DISULFID 2039 2073 BY SIMILARITY.
FT DISULFID 2075 2109 BY SIMILARITY.
FT DISULFID 2111 2145 BY SIMILARITY.
FT DISULFID 2147 2181 BY SIMILARITY.
FT DISULFID 2183 2217 BY SIMILARITY.
FT DISULFID 2219 2253 BY SIMILARITY.
FT DISULFID 2255 2289 BY SIMILARITY.
FT DISULFID 2291 2325 BY SIMILARITY.
FT DISULFID 2327 2361 BY SIMILARITY.
FT DISULFID 2363 2397 BY SIMILARITY.
FT DISULFID 2399 2433 BY SIMILARITY.
FT DISULFID 2435 2469 BY SIMILARITY.
FT DISULFID 2471 2505 BY SIMILARITY.
FT DISULFID 2507 2541 BY SIMILARITY.
FT DISULFID 2543 2577 BY SIMILARITY.
FT DISULFID 2579 2613 BY SIMILARITY.
FT DISULFID 2615 2649 BY SIMILARITY.
FT DISULFID 2651 2685 BY SIMILARITY.
FT DISULFID 2687 2721 BY SIMILARITY.
FT DISULFID 2723 2757 BY SIMILARITY.
FT DISULFID 2759 2793 BY SIMILARITY.
FT DISULFID 2795 2829 BY SIMILARITY.
FT DISULFID 2831 2865 BY SIMILARITY.
FT DISULFID 2867 2901 BY SIMILARITY.
FT DISULFID 2903 2937 BY SIMILARITY.
FT DISULFID 2939 2973 BY SIMILARITY.
FT DISULFID 2975 3009 BY SIMILARITY.
FT DISULFID 3011 3045 BY SIMILARITY.
FT DISULFID 3047 3081 BY SIMILARITY.
FT DISULFID 3083 3117 BY SIMILARITY.
FT DISULFID 3119 3153 BY SIMILARITY.
FT DISULFID 3155 3189 BY SIMILARITY.
FT DISULFID 3191 3225 BY SIMILARITY.
FT DISULFID 3227 3261 BY SIMILARITY.
FT DISULFID 3263 3297 BY SIMILARITY.
FT DISULFID 3299 3333 BY SIMILARITY.
FT DISULFID 3335 3369 BY SIMILARITY.
FT DISULFID 3371 3405 BY SIMILARITY.
FT DISULFID 3407 3441 BY SIMILARITY.
FT DISULFID 3443 3477 BY SIMILARITY.
FT DISULFID 3479 3513 BY SIMILARITY.
FT DISULFID 3515 3549 BY SIMILARITY.
FT DISULFID 3551 3585 BY SIMILARITY.
FT DISULFID 3587 3621 BY SIMILARITY.
FT DISULFID 3623 3657 BY SIMILARITY.
FT DISULFID 3659 3693 BY SIMILARITY.
FT DISULFID 3695 3729 BY SIMILARITY.
FT DISULFID 3731 3765 BY SIMILARITY.
FT DISULFID 3767 3801 BY SIMILARITY.
FT DISULFID 3803 3837 BY SIMILARITY.
FT DISULFID 3839 3873 BY SIMILARITY.
FT DISULFID 3875 3909 BY SIMILARITY.
FT DISULFID 3911 3945 BY SIMILARITY.
FT DISULFID 3947 3981 BY SIMILARITY.
FT DISULFID 3983 4017 BY SIMILARITY.
FT DISULFID 4019 4053 BY SIMILARITY.
FT DISULFID 4055 4089 BY SIMILARITY.
FT DISULFID 4091 4125 BY SIMILARITY.
FT DISULFID 4127 4161 BY SIMILARITY.
FT DISULFID 4163 4197 BY SIMILARITY.
FT DISULFID 4199 4233 BY SIMILARITY.
FT DISULFID 4235 4269 BY SIMILARITY.
FT DISULFID 4271 4305 BY SIMILARITY.
FT DISULFID 4307 4341 BY SIMILARITY.
FT DISULFID 4343 4377 BY SIMILARITY.
FT DISULFID 4379 4413 BY SIMILARITY.
FT DISULFID 4415 4449 BY SIMILARITY.
FT DISULFID 4451 4485 BY SIMILARITY.
FT DISULFID 4487 4521 BY SIMILARITY.
FT DISULFID 4523 4557 BY SIMILARITY.
FT DISULFID 4559 4593 BY SIMILARITY.
FT DISULFID 4595 4629 BY SIMILARITY.
FT DISULFID 4631 4665 BY SIMILARITY.
FT DISULFID 4667 4701 BY SIMILARITY.
FT DISULFID 4703 4737 BY SIMILARITY.
FT DISULFID 4739 4773 BY SIMILARITY.
FT DISULFID 4775 4809 BY SIMILARITY.
FT DISULFID 4811 4845 BY SIMILARITY.
FT DISULFID 4847 4881 BY SIMILARITY.
FT DISULFID 4883 4917 BY SIMILARITY.
FT DISULFID 4919 4953 BY SIMILARITY.
FT DISULFID 4955 4989 BY SIMILARITY.
FT DISULFID 4991 5025 BY SIMILARITY.
FT DISULFID 5027 5061 BY SIMILARITY.
FT DISULFID 5063 5097 BY SIMILARITY.
FT DISULFID 5099 5133 BY SIMILARITY.
FT DISULFID 5135 5169 BY SIMILARITY.
FT DISULFID 5171 5205 BY SIMILARITY.
FT DISULFID 5207 5241 BY SIMILARITY.
FT DISULFID 5243 5277 BY SIMILARITY.
FT DISULFID 5279 5313 BY SIMILARITY.
FT DISULFID 5315 5349 BY SIMILARITY.
FT DISULFID 5351 5385 BY SIMILARITY.
FT DISULFID 5387 5421 BY SIMILARITY.
FT DISULFID 5423 5457 BY SIMILARITY.
FT DISULFID 5459 5493 BY SIMILARITY.
FT DISULFID 5495 5529 BY SIMILARITY.
FT DISULFID 5531 5565 BY SIMILARITY.
FT DISULFID 5567 5601 BY SIMILARITY.
FT DISULFID 5603 5637 BY SIMILARITY.
FT DISULFID 5639 5673 BY SIMILARITY.
FT DISULFID 5675 5709 BY SIMILARITY.
FT DISULFID 5711 5745 BY SIMILARITY.
FT DISULFID 5747 5781 BY SIMILARITY.
FT DISULFID 5783 5817 BY SIMILARITY.
FT DISULFID 5819 5853 BY SIMILARITY.
FT DISULFID 5855 5889 BY SIMILARITY.
FT DISULFID 5891 5925 BY SIMILARITY.
FT DISULFID 5927 5961 BY SIMILARITY.
FT DISULFID 5963 5997 BY SIMILARITY.
FT DISULFID 5999 6033 BY SIMILARITY.
FT DISULFID 6035 6069 BY SIMILARITY.
FT DISULFID 6071 6105 BY SIMILARITY.
FT DISULFID 6107 6141 BY SIMILARITY.
FT DISULFID 6143 6177 BY SIMILARITY.
FT DISULFID 6179 6213 BY SIMILARITY.
FT DISULFID 6215 6249 BY SIMILARITY.
FT DISULFID 6251 6285 BY SIMILARITY.
FT DISULFID 6287 6321 BY SIMILARITY.
FT DISULFID 6323 6357 BY SIMILARITY.
FT DISULFID 6359 6393 BY SIMILARITY.
FT DISULFID 6395 6429 BY SIMILARITY.
FT DISULFID 6431 6465 BY SIMILARITY.
FT DISULFID 6467 6501 BY SIMILARITY.
FT DISULFID 6503 6537 BY SIMILARITY.
FT DISULFID 6539 6573 BY SIMILARITY.
FT DISULFID 6575 6609 BY SIMILARITY.
FT DISULFID 6611 6645 BY SIMILARITY.
FT DISULFID 6647 6681 BY SIMILARITY.
FT DISULFID 6683 6717 BY SIMILARITY.
FT DISULFID 6719 6753 BY SIMILARITY.
FT DISULFID 6755 6789 BY SIMILARITY.
FT DISULFID 6791 6825 BY SIMILARITY.
FT DISULFID 6827 6861 BY SIMILARITY.
FT DISULFID 6863 6897 BY SIMILARITY.
FT DISULFID 6899 6933 BY SIMILARITY.
FT DISULFID 6935 6969 BY SIMILARITY.
FT DISULFID 6971 7005 BY SIMILARITY.
FT DISULFID 7007 7041 BY SIMILARITY.
FT DISULFID 7043 7077 BY SIMILARITY.
FT DISULFID 7079 7113 BY SIMILARITY.
FT DISULFID 7115 7149 BY SIMILARITY.
FT DISULFID 7151 7185 BY SIMILARITY.
FT DISULFID 7187 7221 BY SIMILARITY.
FT DISULFID 7223 7257 BY SIMILARITY.
FT DISULFID 7259 7293 BY SIMILARITY.
FT DISULFID 7295 7329 BY SIMILARITY.
FT DISULFID 7331 7365 BY SIMILARITY.
FT DISULFID 7367 7401 BY SIMILARITY.
FT DISULFID 7403 7437 BY SIMILARITY.
FT DISULFID 7439 7473 BY SIMILARITY.
FT DISULFID 7475 7509 BY SIMILARITY.
FT DISULFID 7511 7545 BY SIMILARITY.
FT DISULFID 7547 7581 BY SIMILARITY.
FT DISULFID 7583 7617 BY SIMILARITY.
FT DISULFID 7619 7653 BY SIMILARITY.
FT DISULFID 7655 7689 BY SIMILARITY.
FT DISULFID 7691 7725 BY SIMILARITY.
FT DISULFID 7727 7761 BY SIMILARITY.
FT DISULFID 7763 7797 BY SIMILARITY.
FT DISULFID 7799 7833 BY SIMILARITY.
FT DISULFID 7835 7869 BY SIMILARITY.
FT DISULFID 7871 7905 BY SIMILARITY.
FT DISULFID 7907 7941 BY SIMILARITY.
FT DISULFID 7943 7977 BY SIMILARITY.
FT DISULFID 7979 8013 BY SIMILARITY.
FT DISULFID 8015 8049 BY SIMILARITY.
FT DISULFID 8051 8085 BY SIMILARITY.
FT DISULFID 8087 8121 BY SIMILARITY.
FT DISULFID 8123 8157 BY SIMILARITY.
FT DISULFID 8159 8193 BY SIMILARITY.
FT DISULFID 8195 8229 BY SIMILARITY.
FT DISULFID 8231 8265 BY SIMILARITY.
FT DISULFID 8267 8301 BY SIMILARITY.
FT DISULFID 8303 8337 BY SIMILARITY.
FT DISULFID 8339 8373 BY SIMILARITY.
FT DISULFID 8375 8409 BY SIMILARITY.
FT DISULFID 8411 8445 BY SIMILARITY.
FT DISULFID 8447 8481 BY SIMILARITY.
FT DISULFID 8483 8517 BY SIMILARITY.
FT DISULFID 8519 8553 BY SIMILARITY.
FT DISULFID 8555 8589 BY SIMILARITY.
FT DISULFID 8591 8625 BY SIMILARITY.
FT DISULFID 8627 8661 BY SIMILARITY.
FT DISULFID 8663 8697 BY SIMILARITY.
FT DISULFID 8699 8733 BY SIMILARITY.
FT DISULFID 8735 8769 BY SIMILARITY.
FT DISULFID 8771 8805 BY SIMILARITY.
FT DISULFID 8807 8841 BY SIMILARITY.
FT DISULFID 8843 8877 BY SIMILARITY.
FT DISULFID 8879 8913 BY SIMILARITY.
FT DISULFID 8915 8949 BY SIMILARITY.
FT DISULFID 8951 8985 BY SIMILARITY.
FT DISULFID 8987 9021 BY SIMILARITY.
FT DISULFID 9023 9057 BY SIMILARITY.
FT DISULFID 9059 9093 BY SIMILARITY.
FT DISULFID 9095 9129 BY SIMILARITY.
FT DISULFID 9131 9165 BY SIMILARITY.
FT DISULFID 9167 9201 BY SIMILARITY.
FT DISULFID 9203 9237 BY SIMILARITY.
FT DISULFID 9239 9273 BY SIMILARITY.
FT DISULFID 9275 9309 BY SIMILARITY.
FT DISULFID 9311 9345 BY SIMILARITY.
FT DISULFID 9347 9381 BY SIMILARITY.
FT DISULFID 9383 9417 BY SIMILARITY.
FT DISULFID 9419 9453 BY SIMILARITY.
FT DISULFID 9455 9489 BY SIMILARITY.
FT DISULFID 9491 9525 BY SIMILARITY.
FT DISULFID 9527 9561 BY SIMILARITY.
FT DISULFID 9563 9597 BY SIMILARITY.
FT DISULFID 9599 9633 BY SIMILARITY.
FT DISULFID 9635 9669 BY SIMILARITY.
FT DISULFID 9671 9705 BY SIMILARITY.
FT DISULFID 9707 9741 BY SIMILARITY.
FT DISULFID 9743 9777 BY SIMILARITY.
FT DISULFID 9779 9813 BY SIMILARITY.
FT DISULFID 9815 9849 BY SIMILARITY.
FT DISULFID 9851 9885 BY SIMILARITY.
FT DISULFID 9887 9921 BY SIMILARITY.
FT DISULFID 9923 9957 BY SIMILARITY.
FT DISULFID 9959 9993 BY SIMILARITY.
FT DISULFID 9995 10000 BY SIMILARITY.

```

```

DB 1077 ---VQKARPHCLCPGMDGAYCDVLNVSCRAALQKG 1111
RESULT 11
ID IGIR_RAT STANDARD; PRT; 1370 AA.
AC P24062;
DT 01-MAR-1992 (Rel. 21, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Insulin-like growth factor I receptor precursor (EC 2.7.1.112).
GN IGIR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxId=10116;
[1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RX MEDLINE=95277910; PubMed=7758167;
RA Du J., Delafontaine P.;
RT "Inhibition of vascular smooth muscle cell growth through antisense
transcription of a rat insulin-like growth factor I receptor cDNA.";
RL Circ. Res. 76:963-972(1995).
[2]
RP SEQUENCE OF 1-364 FROM N.A.
RC STRAIN=Sprague-Dawley;
RX MEDLINE=90017496; PubMed=2477843;
RA Werner H., Moloschak M., Adamo M., Shen-Orr Z., Roberts C.T. Jr.,
RA Lerofth D.;
RT "Developmental regulation of the rat insulin-like growth factor I
receptor gene.";
RL Proc. Natl. Acad. Sci. U.S.A. 86:7451-7455(1989).
[3]
RP SEQUENCE OF 913-1017 FROM N.A.
RX MEDLINE=92412145; PubMed=1530648;
RA Kurachi H., Jodo K., Ohta M., Kawasaki T., Itoh N.;
RT "A new member of the insulin receptor family, insulin
receptor-related receptor, is expressed preferentially in the
kidney.";
RL Biochem. Biophys. Res. Commun. 187:934-939(1992).
CC -I- FUNCTION: THIS RECEPTOR BINDS INSULIN-LIKE GROWTH FACTOR I (IGF I)
WITH A HIGH AFFINITY AND IGF II WITH A LOWER AFFINITY. IT HAS A
TYROSINE-PROTEIN KINASE ACTIVITY.
CC -I- CATALYTIC ACTIVITY: ATP + a protein tyrosine = ADP + protein
tyrosine phosphate.
CC -I- SUBUNIT: Tetramer of 2 alpha and 2 beta chains linked by disulfide
bonds. The alpha chains contribute to the formation of the ligand-
binding domain, while the beta chain carries the kinase domain.
CC -I- SUBCELLULAR LOCATION: Type I membrane protein.
CC -I- SIMILARITY: BELONGS TO THE TYR FAMILY OF PROTEIN KINASES. INSULIN
RECEPTOR SUBFAMILY.
CC -I- SIMILARITY: CONTAINS 2 FIBRONECTIN TYPE III-LIKE DOMAINS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
CC EMBL: L29232; AAA4192.1; -
CC EMBL: M27293; AAA41384.1; -
CC PIR: A33837; A33837.
CC HSSP: P06213; 1HKR.
DR InterPro: IPR000494; EGF_L1_domain.
DR InterPro: IPR000719; Euk_kinase.
DR InterPro: IPR003961; FN_III.
DR InterPro: IPR002174; Furin-like.
DR InterPro: IPR002011; RTkinaseII.
DR InterPro: IPR001245; Tyr_kinase.
DR Pfam: PR00041; fn3; 2.

```

```

DR Pfam: PF00069; Pkinase; 1.
DR Pfam: PF00757; Furin-like; 1.
DR Pfam: PF01030; Recep. I domain; 2.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Euk_Pkinase; 1.
DR SMART; SM00060; FN3; 3.
DR SMART; SM00261; FU; 1.
DR SMART; SM00219; TYKIC; 1.
DR PROSITE; PS00107; PROTEIN KINASE ATP; 1.
DR PROSITE; PS00109; PROTEIN KINASE_TYR; 1.
DR PROSITE; PS00239; RECEPTOR_TYR_KIN_II; 1.
DR PROSITE; PS00011; PROTEIN KINASE_DOM; 1.
DR Transferrase; Tyrosine-protein kinase; Receptor; Transmembrane;
KM Glycoprotein; ATP-binding; Phosphorylation; Repeat; Signal.
FT SIGNAL 1 30
FT CHAIN 31 741
FT CHAIN 742 1370
FT DOMAIN 742 936
FT TRANSMEM 937 960
FT DOMAIN 961 1370
FT DOMAIN 608 829
FT DOMAIN 830 929
FT DOMAIN 1000 1275
FT NP_BIND 1006 1014
FT BINDING 1034 1034
FT ACT_SITE 1136 1136
FT DISULFID 215 224
FT DISULFID 219 230
FT DISULFID 231 239
FT DISULFID 235 248
FT DISULFID 251 260
FT DISULFID 264 276
FT DISULFID 282 303
FT DISULFID 307 321
FT DISULFID 324 328
FT DISULFID 51 51
FT CARBOHYD 102 102
FT CARBOHYD 135 135
FT CARBOHYD 245 245
FT CARBOHYD 314 314
FT CARBOHYD 418 418
FT CARBOHYD 439 439
FT CARBOHYD 535 535
FT CARBOHYD 608 608
FT CARBOHYD 623 623
FT CARBOHYD 641 641
FT CARBOHYD 748 748
FT CARBOHYD 757 757
FT CARBOHYD 765 765
FT CARBOHYD 901 901
FT CARBOHYD 914 914
FT MOD_RES 1166 1166
FT CONFLICT 985 986
SO SEQUENCE 1370 AA; 155395 MW; A5946897A1CB145 CRC64;

Query Match 9 83; Score 149; DB 1; Length 1370;
Best Local Similarity 26.7%; Pred. No. 0.0024;
Matches 47; Conservative 17; Mismatches 46; Indels 66; Gaps 11;

```

```

RESULT 12
NOTCH_XENLA STANDARD; PRT; 2524 AA.
AC P21783;
DT 01-MAY-1991 (rel. 18, Created)
DT 01-OCT-1996 (rel. 34, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Neurogenic locus notch protein homolog (XOTCH protein).
GN XOTCH.
OS Xenopus laevis (African clawed frog).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae; Pipidae;
OC Xenopodidae; Xenopus.
OX NCBI_TaxID=8355;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90385285; Pubmed=2402639;
RA Coffman C., Harris W., Kintner C.;
RT "Notch, the Xenopus homolog of Drosophila notch.";
RL Science 249:1438-1441(1990).
RN [2]
RP REVISIONS TO 1759-1782.
RA Kintner C.;
RL Submitted (JUN-1996) to the EMBL/Genbank/DBJ databases.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED ALMOST UNIFORMLY IN EARLY EMBRYOS.
CC -1- SIMILARITY: BELONGS TO THE NOTCH FAMILY.
CC -1- SIMILARITY: CONTAINS 36 EGF-LIKE DOMAINS.
CC -1- SIMILARITY: CONTAINS 3 LIN/NOTCH REPEATS.
CC -1- SIMILARITY: CONTAINS 6 ANK REPEATS.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation.
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; M33874; AAB02039.1; -.
DR PIR; A33844; A33844.
DR HSP; P00740; 1EDM.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR000152; Asx_hydroxyl.
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR000742; EGF_2.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR001438; EGF_II.
DR InterPro; IPR000800; Notch.
DR Pfam; PF00008; EGF; 36.
DR Pfam; PF00023; ank; 6.
DR Pfam; PF00060; notch; 3.
DR PRINTS; PR00010; EGFBL00D.
DR PRINTS; PR01452; NOTCH.
DR SMART; SM00248; ANK; 5.
DR SMART; SM00179; EGF_CA; 23.
DR SMART; SM00001; EGF_Like; 11.
DR SMART; SM00004; NL; 2.
DR PROSITE; PS00088; ANK_REPEAT; 4.
DR PROSITE; PS00297; ANK_REPEAT_REGION; 1.
DR PROSITE; PS00022; ASX_HYDROXYL; 23.
DR PROSITE; PS00010; EGF_1; 34.
DR PROSITE; PS01186; EGF_2; 29.
DR PROSITE; PS01187; EGF_CA; 21.
DR Differentiation; Neurogenesis; Repeat; ANK repeat; EGF-like domain;
KW Transmembrane; Signal; Glycoprotein.
FT SIGNAL 1 19
FT CHAIN 20 2524
FT DOMAIN 20 1728
FT TRANSMEM 1729 1750
FT DOMAIN 1751 2524
FT DOMAIN 20 57
FT DOMAIN 20 57
EGF-Like 1.

```

FT	DOMAIN	58	99	EGF-LIKE 2.
FT	DOMAIN	102	140	EGF-LIKE 3.
FT	DOMAIN	141	177	EGF-LIKE 4.
FT	DOMAIN	179	215	EGF-LIKE 5.
FT	DOMAIN	217	254	EGF-LIKE 6.
FT	DOMAIN	256	292	EGF-LIKE 7.
FT	DOMAIN	294	332	EGF-LIKE 8.
FT	DOMAIN	334	370	EGF-LIKE 9.
FT	DOMAIN	371	409	EGF-LIKE 10.
FT	DOMAIN	411	449	EGF-LIKE 11.
FT	DOMAIN	451	487	EGF-LIKE 12.
FT	DOMAIN	489	525	EGF-LIKE 13.
FT	DOMAIN	527	563	EGF-LIKE 14.
FT	DOMAIN	565	600	EGF-LIKE 15.
FT	DOMAIN	602	638	EGF-LIKE 16.
FT	DOMAIN	640	675	EGF-LIKE 17.
FT	DOMAIN	677	713	EGF-LIKE 18.
FT	DOMAIN	715	750	EGF-LIKE 19.
FT	DOMAIN	752	788	EGF-LIKE 20.
FT	DOMAIN	790	826	EGF-LIKE 21.
FT	DOMAIN	828	866	EGF-LIKE 22.
FT	DOMAIN	868	904	EGF-LIKE 23.
FT	DOMAIN	906	942	EGF-LIKE 24.
FT	DOMAIN	944	980	EGF-LIKE 25.
FT	DOMAIN	982	1018	EGF-LIKE 26.
FT	DOMAIN	1020	1056	EGF-LIKE 27.
FT	DOMAIN	1058	1094	EGF-LIKE 28.
FT	DOMAIN	1096	1142	EGF-LIKE 29.
FT	DOMAIN	1144	1180	EGF-LIKE 30.
FT	DOMAIN	1182	1218	EGF-LIKE 31.
FT	DOMAIN	1220	1264	EGF-LIKE 32.
FT	DOMAIN	1266	1304	EGF-LIKE 33.
FT	DOMAIN	1306	1346	EGF-LIKE 34.
FT	DOMAIN	1347	1383	EGF-LIKE 35.
FT	DOMAIN	1386	1424	EGF-LIKE 36.
FT	REPEAT	1441	1478	LIN/NOTCH 1.
FT	REPEAT	1479	1520	LIN/NOTCH 2.
FT	REPEAT	1521	1560	LIN/NOTCH 3.
FT	REPEAT	1576	1619	ANK 1.
FT	REPEAT	1624	1663	ANK 2.
FT	REPEAT	1676	1719	ANK 3.
FT	REPEAT	1724	1767	ANK 4.
FT	REPEAT	1776	1819	ANK 5.
FT	REPEAT	1824	1866	ANK 6.
FT	DISULFID	22	35	BY SIMILARITY.
FT	DISULFID	29	45	BY SIMILARITY.
FT	DISULFID	47	56	BY SIMILARITY.
FT	DISULFID	62	74	BY SIMILARITY.
FT	DISULFID	68	87	BY SIMILARITY.
FT	DISULFID	89	98	BY SIMILARITY.
FT	DISULFID	106	117	BY SIMILARITY.
FT	DISULFID	111	128	BY SIMILARITY.
FT	DISULFID	130	139	BY SIMILARITY.
FT	DISULFID	145	156	BY SIMILARITY.
FT	DISULFID	150	165	BY SIMILARITY.
FT	DISULFID	167	176	BY SIMILARITY.
FT	DISULFID	183	194	BY SIMILARITY.
FT	DISULFID	188	203	BY SIMILARITY.
FT	DISULFID	205	214	BY SIMILARITY.
FT	DISULFID	221	232	BY SIMILARITY.
FT	DISULFID	226	242	BY SIMILARITY.
FT	DISULFID	244	253	BY SIMILARITY.
FT	DISULFID	260	271	BY SIMILARITY.
FT	DISULFID	265	280	BY SIMILARITY.
FT	DISULFID	282	291	BY SIMILARITY.
FT	DISULFID	298	311	BY SIMILARITY.
FT	DISULFID	305	320	BY SIMILARITY.
FT	DISULFID	322	331	BY SIMILARITY.
FT	DISULFID	338	349	BY SIMILARITY.
FT	DISULFID	343	358	BY SIMILARITY.
FT	DISULFID	360	369	BY SIMILARITY.
FT	DISULFID	375	386	BY SIMILARITY.
FT	DISULFID	380	397	BY SIMILARITY.

FT	DISULFID	399	408	BY SIMILARITY.
FT	DISULFID	415	428	BY SIMILARITY.
FT	DISULFID	422	437	BY SIMILARITY.
FT	DISULFID	439	448	BY SIMILARITY.
FT	DISULFID	455	466	BY SIMILARITY.
FT	DISULFID	460	475	BY SIMILARITY.
FT	DISULFID	477	486	BY SIMILARITY.
FT	DISULFID	493	504	BY SIMILARITY.
FT	DISULFID	498	513	BY SIMILARITY.
FT	DISULFID	515	524	BY SIMILARITY.
FT	DISULFID	531	542	BY SIMILARITY.
FT	DISULFID	536	551	BY SIMILARITY.
FT	DISULFID	553	562	BY SIMILARITY.
FT	DISULFID	559	579	BY SIMILARITY.
FT	DISULFID	574	588	BY SIMILARITY.
FT	DISULFID	590	599	BY SIMILARITY.
FT	DISULFID	606	617	BY SIMILARITY.
FT	DISULFID	611	626	BY SIMILARITY.
FT	DISULFID	628	637	BY SIMILARITY.
FT	DISULFID	644	654	BY SIMILARITY.
FT	DISULFID	649	663	BY SIMILARITY.
FT	DISULFID	665	674	BY SIMILARITY.
FT	DISULFID	681	692	BY SIMILARITY.
FT	DISULFID	686	701	BY SIMILARITY.
FT	DISULFID	703	712	BY SIMILARITY.
FT	DISULFID	719	729	BY SIMILARITY.
FT	DISULFID	724	738	BY SIMILARITY.
FT	DISULFID	740	749	BY SIMILARITY.
FT	DISULFID	756	767	BY SIMILARITY.
FT	DISULFID	761	776	BY SIMILARITY.
FT	DISULFID	778	787	BY SIMILARITY.
FT	DISULFID	794	805	BY SIMILARITY.
FT	DISULFID	799	814	BY SIMILARITY.
FT	DISULFID	816	825	BY SIMILARITY.
FT	DISULFID	832	843	BY SIMILARITY.
FT	DISULFID	837	854	BY SIMILARITY.
FT	DISULFID	856	865	BY SIMILARITY.
FT	DISULFID	872	883	BY SIMILARITY.
FT	DISULFID	877	892	BY SIMILARITY.
FT	DISULFID	894	903	BY SIMILARITY.
FT	DISULFID	910	921	BY SIMILARITY.
FT	DISULFID	915	930	BY SIMILARITY.
FT	DISULFID	932	941	BY SIMILARITY.
FT	DISULFID	986	997	BY SIMILARITY.
FT	DISULFID	991	1006	BY SIMILARITY.
FT	DISULFID	1008	1017	BY SIMILARITY.
FT	DISULFID	1024	1035	BY SIMILARITY.
FT	DISULFID	1029	1044	BY SIMILARITY.
FT	DISULFID	1046	1055	BY SIMILARITY.
FT	DISULFID	1062	1073	BY SIMILARITY.
FT	DISULFID	1067	1082	BY SIMILARITY.
FT	DISULFID	1084	1093	BY SIMILARITY.
FT	DISULFID	1100	1121	BY SIMILARITY.
FT	DISULFID	1115	1130	BY SIMILARITY.
FT	DISULFID	1132	1141	BY SIMILARITY.
FT	DISULFID	1148	1159	BY SIMILARITY.
FT	DISULFID	1153	1168	BY SIMILARITY.

Query Match 9.8%; Score 148.5; DB 1; Length 2524;
Best Local Similarity 22.5%; Pred. No. 0.0043;
Matches 73; Conservative 32; Mismatches 102; Indels 117; Gaps 19;

Qy	34	MHPNVSQCGGCGATCSYNGCSCKRPLFALERIGM-----KQIGV 76
Db	869	MNECVNRPERNG-ATCQNTNGSYKCKCKPGYTGRCNEWDIDDCCQPNCHNGSCSDGINM 927
Qy	77	CLASSCPGYYGTR-YPDINKCTK--CK--ADCDTCFNKPFCKKSGFY-LHL-----123
Db	928	FFCNCPAGRGKCEEDINECASPNCKNGANCTDCVNSYCTC-CQPFSGIHESNTPPC 986
Qy	124	-----GKCLD-----NCEGLEAN--NHTM-ECVSI-----146
Db	987	TESSCNGSTCIDGINPTCCCPGFTGSGYCOHDINECDSPCLNGSTCODSYGYTKCTC 1046

Query Match	Best Local Similarity	9.7%	Score 147.5; DB 1;	Length 2471;
FT DOMAIN	795	831	EGF-LIKE 21, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	833	871	EGF-LIKE 22, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	873	909	EGF-LIKE 23, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	911	947	EGF-LIKE 24, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	949	985	EGF-LIKE 25, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	987	1023	EGF-LIKE 26, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1025	1061	EGF-LIKE 27, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1063	1099	EGF-LIKE 28, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1101	1147	EGF-LIKE 29, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1149	1185	EGF-LIKE 30, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1187	1223	EGF-LIKE 31, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1225	1262	EGF-LIKE 32, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1264	1302	EGF-LIKE 33, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1304	1343	EGF-LIKE 34, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1374	1412	EGF-LIKE 35, CALCITON-BINDING (POTENTIAL).	
FT DOMAIN	1445	1484	POLY-ALA.	
FT DOMAIN	1494	1533	POLY-LEU.	
FT DOMAIN	1594	1633	POLY-SER.	
FT DOMAIN	1646	1685	POLY-GLY.	
FT DOMAIN	1746	1785	POLY-GLY.	
FT DOMAIN	1827	1866	LIN/NOTCH 1.	
FT REPEAT	1871	1910	LIN/NOTCH 2.	
FT REPEAT	1905	1944	ANK 1.	
FT REPEAT	1939	1978	ANK 2.	
FT REPEAT	1972	2011	ANK 3.	
FT REPEAT	1976	2015	ANK 4.	
FT REPEAT	2009	2048	ANK 5.	
FT REPEAT	2038	2077	ANK 6.	
FT DISULFID	28	67	BY SIMILARITY.	
FT DISULFID	35	74	BY SIMILARITY.	
FT DISULFID	53	92	BY SIMILARITY.	
FT DISULFID	68	107	BY SIMILARITY.	
FT DISULFID	73	112	BY SIMILARITY.	
FT DISULFID	92	131	BY SIMILARITY.	
FT DISULFID	109	148	BY SIMILARITY.	
FT DISULFID	115	154	BY SIMILARITY.	
FT DISULFID	133	172	BY SIMILARITY.	
FT DISULFID	148	188	BY SIMILARITY.	
FT DISULFID	153	193	BY SIMILARITY.	
FT DISULFID	170	210	BY SIMILARITY.	
FT DISULFID	186	226	BY SIMILARITY.	
FT DISULFID	192	232	BY SIMILARITY.	
FT DISULFID	209	249	BY SIMILARITY.	
FT DISULFID	225	265	BY SIMILARITY.	
FT DISULFID	230	270	BY SIMILARITY.	
FT DISULFID	248	288	BY SIMILARITY.	
FT DISULFID	264	304	BY SIMILARITY.	
FT DISULFID	269	309	BY SIMILARITY.	
FT DISULFID	286	326	BY SIMILARITY.	
FT DISULFID	302	342	BY SIMILARITY.	
FT DISULFID	309	349	BY SIMILARITY.	
FT DISULFID	326	366	BY SIMILARITY.	
FT DISULFID	342	382	BY SIMILARITY.	
FT DISULFID	347	387	BY SIMILARITY.	
FT DISULFID	364	404	BY SIMILARITY.	
FT DISULFID	379	419	BY SIMILARITY.	
FT DISULFID	384	424	BY SIMILARITY.	
FT DISULFID	403	443	BY SIMILARITY.	
FT DISULFID	419	459	BY SIMILARITY.	
FT DISULFID	427	467	BY SIMILARITY.	
FT DISULFID	444	484	BY SIMILARITY.	
FT DISULFID	460	500	BY SIMILARITY.	
FT DISULFID	471	511	BY SIMILARITY.	
FT DISULFID	480	520	BY SIMILARITY.	
FT DISULFID	491	531	BY SIMILARITY.	
FT DISULFID	498	538	BY SIMILARITY.	
FT DISULFID	509	549	BY SIMILARITY.	
FT DISULFID	520	560	BY SIMILARITY.	
FT DISULFID	536	576	BY SIMILARITY.	
FT DISULFID	541	581	BY SIMILARITY.	
FT DISULFID	558	598	BY SIMILARITY.	
FT DISULFID	574	614	BY SIMILARITY.	

Matches 54; Conservative 23; Mismatches 52; Indels 89; Gaps 13;

31 ORMHVNVGSGGACATCDYNGCLSCXPRFLPFLERIGMIOVICSSPSGYTRY 90
 947 QTDNCELEPCKNG-GTCSYVNSYTC-----TPAGFHVHC 984
 91 P-DINKCTCKKADCTCFENKNC-----TKCKSGFYHLGK 125
 985 ENNIDECTE-----SSCNGCGTCVDGINSFSCLPVGTGPECLHDINECSNPLNSGT 1039
 126 CLD-----NCPGLEAMNHMECVSIH-CEVSEMNPMSPCTCKGTCGFKGTETRV 177
 1040 CVDGLGTYCTCPGLGYTKN-----CQTLVNLG-----SP-SCKKNG-TCA----- 1079
 178 REIIQHPKSAKMLCPPTNETRKTGVRKRC-----OKG 210
 1080 ----OEKARPRCLCPGWDGAYCDVLNWSCKAALOKG 1113

RESULT 14
 BL14_CABEL STANDARD; PRT; 943 AA.
 AC PS1559; 044762; 044763; 044764; 044765; 044766;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Endoprotease bli-4 precursor (EC 3.4.21.-) (Blisterase) (Blistered cuticle protein 4).
 CN BLI-4 OR K04P10.4.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditicoidea;
 OC Rhabditidae; Pelodierinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 [1]
 SEQUENCE FROM N.A. (ISOFORMS A; B; C AND D), AND FUNCTION.
 RP STRAIN-Bristol N2;
 RX MEDLINE=95293228; PubMed=7774813;
 RA Thacker C, Peters K W, Gray M, Rose A M;
 RT "The bli-4 locus of Caenorhabditis elegans encodes structurally distinct hex2/subtilisin-like endoproteases essential for early development and adult morphology.";
 RT Genes Dev. 9:956-971(1995).
 RL [2]
 SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
 RP STRAIN-Bristol N2;
 RC Lattelle P, Mamsley P;
 RA Submitted (Jan-1998) to the EMBL/Genbank/DBJ databases.
 CC -I- FUNCTION: Isoform A is required for normal production or maintenance of the adult cuticle. There is functional redundancy between the isoforms. Vital for embryonic and larval development.
 CC -I- TISSUE SPECIFICITY: In larvae and adults, expressed in all hypodermal cells, vulva and ventral nerve cords.
 CC -I- DEVELOPMENTAL STAGE: Expression starts at embryo two-fold stage through to adults.
 CC -I- ALTERNATIVE PRODUCTS: 5 isoforms: A/A, B/B, C/C, D/D (shown here) and e; may be produced by alternative splicing.
 CC -I- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8, FURIN SUBFAMILY.
 CC -I- CAUTION: Rel.1 sequence differs from that shown due to a frameshift in position 651.
 CC -I- CAUTION: Ref.2 (AA96754/AA96757) sequence differs from that shown due to erroneous gene model prediction.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
 CC
 DR EMBL; I29438; AAA98750.1; ALT_FRAME.
 DR EMBL; I29439; AAA98751.1; ALT_FRAME.
 DR EMBL; I29440; AAA98752.1; ALT_FRAME.

RESULT 15	WIFI_MOUSE	STANDARD;	PRT;	379 AA.
AC	09WUAI.			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	16-OCT-2001 (Rel. 40, Last annotation update)			
DE	Mnt inhibitory factor 1 precursor (WIF-1).			
GN	WIF1.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Euteleia; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_Taxid=10090;			
RN	(1)			
RX	SEQUENCE FROM N.A.			
RA	MEDLINE=99215557; PubMed=10201374;			
RA	Hsieh J.-C., Kodjabachian L., Rebbeck M.L., Rattner A.,			
RA	Smallwood P.M., Samos C.H., Nusse R., David I.B., Nahans J.;			
RT	"A new secreted protein that binds to Mnt proteins and inhibits their			
RT	activities".			
RT	Nature 398:431-436(1999).			
CC	- FUNCTION: BINDS TO MNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY			
CC	BE INVOLVED IN MESODERM SEGMENTATION.			
CC	- SUBCELLULAR LOCATION: Secreted.			
CC	- TISSUE SPECIFICITY: EXPRESSION HIGHEST IN HEART AND LUNG. LOWER IN			
CC	BRAIN AND EYE.			
CC	- SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL, AF122923; AAD25403.1; -			
DR	HSSP, P00740.1EDM.			
DR	MGD: MGI:1344332; Wif1.			
DR	InterPro, IPR0000561; EGF-like			
DR	InterPro, IPR002049; Laminin_EGF			
DR	InterPro, IPR003306; WIF.			
DR	Pfam, PF00008; EGF, 5.			
DR	Pfam, PF02019; WIF, 1.			
DR	PRINTS, PR00011; EGF_LAMININ.			
DR	SMART, SMO0181; EGF, 5.			
DR	SMART, SMO0469; WIF, 1.			
DR	PROSITE, PS00022; EGF_1, 5.			
DR	PROSITE, PS0186; EGF_2, 4.			
KM	Repeat, EGF-like domain; Signal; Developmental protein.			
FT	SIGNAL	1	28	POTENTIAL.
FT	CHAIN	29	379	MNT INHIBITORY FACTOR 1.
FT	DOMAIN	177	208	EGF-LIKE 1.
FT	DOMAIN	209	240	EGF-LIKE 2.
FT	DOMAIN	241	272	EGF-LIKE 3.
FT	DOMAIN	273	304	EGF-LIKE 4.
FT	DOMAIN	305	336	EGF-LIKE 5.
FT	DISULFID	177	186	POTENTIAL.
FT	DISULFID	182	192	POTENTIAL.
FT	DISULFID	198	200	POTENTIAL.
FT	DISULFID	209	218	POTENTIAL.
FT	DISULFID	214	224	POTENTIAL.
FT	DISULFID	230	232	POTENTIAL.
FT	DISULFID	241	250	POTENTIAL.
FT	DISULFID	246	256	POTENTIAL.
FT	DISULFID	262	264	POTENTIAL.
FT	DISULFID	273	282	POTENTIAL.
FT	DISULFID	278	288	POTENTIAL.
FT	DISULFID	294	296	POTENTIAL.
FT	DISULFID	305	314	POTENTIAL.
FT	DISULFID	310	320	POTENTIAL.
FT	DISULFID	326	328	POTENTIAL.

FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL)
SQ SEQUENCE 379 AA; 41590 MM; E3765F2642B2BC9A CRC64;

Query Match 9.5%; Score 143.5; DB 1; Length 379;
Best Local Similarity 24.3%; Pred. No. 0.0019;
Matches 60; Conservative 24; Mismatches 68; Indels 95; Gaps 17;

Qy 42 CCGGCAT---CSDYNGC-----LSCKPRLFFALERIGMKQIGVCLSS---C 81
Db 182 CPGGCRNGCFENRRVCECPDGFYGHCEKALCIPRCMNG-----GLCVTPGFCIC 232
Qy 82 PSGYGTVPDINKCTKC-KADCD-TCFNKNPCTCKSGFYHLGKCLDNCPEGLEANNH 139
Db 233 PPGFYG-----VNCDRANCSITCFNGGTC-----FY--PGKCI--CPGLEGE-- 271
Qy 140 TMECVSIVHCEVSEMEPMSPCTKKGKTCGFKGTETRVREIIQHPSA-KGNLCP----- 192
Db 272 -----QCELSKCP--QPCRNKGKCI-----KSKCKCPKGYGDLCKRVCEP 312
Qy 193 -----PTNETRKTCTVQRKKCKGKGRER-----KPKPKGSKKALPDKSLE 240
Db 313 GCGAHGTCHEPNKC-----QCREGWHGRHCHNRRYGASIMHAPRPAGLEHRTTSLKAE 367
Qy 241 SSKELPE 247
Db 368 DRDPPE 374

Search completed: May 6, 2003, 14:52:00
Job time : 16 secs

GenCore version 5.1.4.ps.4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 14:50:38 ; Search time 18 Seconds

(Without alignments) 1458.039 Million cell updates/sec

Title: US-09-894-912a-13

Perfect score: 1516

Sequence: 1 MGHRLISMFLINFMEMYI.....QQKKRVQKQSVSVTVH 273

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 263224 seqs, 96134422 residues

263224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: PIR 73:*

1: PIR1:*

2: PIR2:*

3: PIR3:*

4: PIR4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	192	12.7	1299	2 T43251	furin (EC 3.4.21.7
2	186.5	12.3	1548	2 S34583	serine proteinase
3	184	12.1	962	2 UC5571	subtilisin-like pr
4	184	12.1	975	2 UC5570	subtilisin-like pr
5	176.5	11.6	1680	2 A33434	furin (EC 3.4.21.7
6	173.5	11.4	969	1 A39490	subtilisin-like pr
7	168	11.1	915	2 B48225	probable propeptel
8	167.5	11.0	932	2 I52527	PACE4A - mouse (fr
9	167	11.0	915	1 A48225	subtilisin-like pr
10	162	10.7	899	2 G02428	subtilisin-like pr
11	162	10.7	915	2 UC6148	subtilisin-like pr
12	161	10.6	440	2 T24232	hypothetical prote
13	158.5	10.5	379	2 A59180	hypothetical prote
14	156.5	10.3	937	2 I53282	gene PACB4 protein
15	156	10.3	803	2 A47723	F-spondin precursor
16	153.5	10.1	942	2 D87803	protein b1i-4D (lm
17	150.5	9.9	1203	2 A49175	insulin-like growt
18	149	9.8	1371	2 A33837	insulin-like growt
19	148.5	9.8	2524	2 A35844	Xorch protein - Af
20	147.5	9.7	2471	2 A49128	cell-face determin
21	145	9.6	570	2 T37314	probable kexin (EC
22	145	9.6	1620	2 T27283	hypothetical prote
23	143.5	9.5	807	2 A38152	F-spondin - rat
24	142	9.4	378	2 B59180	Wnt inhibitory fac
25	142	9.4	677	2 C42125	trichophozoite cyste
26	139	9.2	1367	1 IGHU01	insulin-like growt
27	139	9.2	2555	2 A40043	protein-tyrosine k
28	138.5	9.1	1369	2 S70713	noct protein homo
29	138	9.1	2531	2 S18188	noct protein homo

30	138	9.1	2531	2 A46019	Notch-1 protein -
31	137	9.0	2437	2 S42612	transmembrane prot
32	136	9.0	540	2 B47417	insulin receptor-r
33	135.5	8.9	1382	1 INHUR	insulin receptor p
34	134.5	8.9	1111	2 T26972	hypothetical prote
35	134.5	8.9	1372	2 A34157	insulin receptor p
36	134.5	8.9	1383	2 A36080	insulin receptor p
37	129.5	8.5	2101	2 S57245	insulin receptor (
38	129.5	8.5	2148	1 A56081	insulin receptor -
39	128	8.4	327	2 A46484	apoptosis-mediatin
40	128	8.4	1268	2 B36502	insulin receptor-r
41	127	8.4	861	2 A48825	Notch homolog Motc
42	126.5	8.3	3871	2 T22812	hypothetical prote
43	125.5	8.3	667	2 A48579	trichophozoite surfac
44	125	8.2	837	2 S43656	furin (EC 3.4.21.7
45	124	8.2	1274	2 T42017	cysteine rich prot

ALIGNMENTS

RESULT 1
T43251
furin (EC 3.4.21.75) - fall armyworm
N:Alternate names: paired basic amino acid cleaving enzyme; proprotein convertase; seri
C:Species: Spodoptera frugiperda (fall armyworm)
C>Date: 11-Jan-2000 #sequence_revision 11-Jan-2000 #text_change 11-Jan-2000
C:Accession: T43251
R:Cieplik, M.; Klenk, H.
Submitted to the EMBL Data Library, January 1996
A:Description: Cloning and functional characterization of furin from Spodoptera frugiperda
A:Reference number: Z22368
A:Accession: T43251
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-1299 <CF>
A:Cross-references: EMBL:Z68888; NID:g1167859; PID:e2196590; PIDN:CAA91116.1
A:Experimental source: Clone Sfurin 6; ovary
C:Function:
A:Description: responsible for the endoproteolytic processing of proproteins with speci
C:Keywords: hydrolase; serine proteinase

Query Match 12.7% Score 192; DB 2; Length 1299;
Best local similarity 27.8%; Pred No. 2.4e-05;
Matches 63; Conservative 26; Mismatches 78; Indels 60; Gaps 12;

OY 37 NVSGGCGCATCS-D-YNGCLSKPRLPALERIGMKOIGVCLSSCPGYYGTRPDINK 95
Db 795 SVGRPCAHKCTCERADGCTSCENHL-----VLHDTGCMASCPSHYET---EDM 843
OY 96 CTRKADDPDCE--NKPCTKCKSGFYHLGKCLDNCPEGLFANNHTMECVSIHCEVSE 153
Db 844 CAKCHESCDTQCGEGVOCYCTHPSTYALDGCYVSCPPAYYADKKRKEC---MRCPVG- 899
OY 154 KNPMSPTCKKGTGCFKRGTRVREIIQHPAKGNLCPP---TNETRKC-TVORRCKOK 209
Db 900 ---CSTCT-----SAFLSCPEWELNKKKGMVGSDDKCSA 933
OY 210 GERGKGRERKKPKN-----GESKE---AIPDSKLSSEKIEP 247
Db 934 GEFAY---DQCKRCNPACSCVGENEGHCLTCEPNNLDQYKCVPE 977

RESULT 2

S34583
serine proteinase (EC 3.4.21.-) PC6B - mouse
C:Species: Mus musculus (house mouse)
C>Date: 02-Dec-1993 #sequence_revision 10-Nov-1995 #text_change 05-Nov-1999
C:Accession: S34583
R:Nakagawa, T.; Murakami, K.; Nakayama, K.
FEBS Lett. 327, 165-171, 1993
A>Title: Identification of an isoform with an extremely large Cys-rich region of PC6, a
A:Reference number: S34583; M0ID:93327934; PMID:835106

A:Accession: A43434
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1680 <ROF>
A:Cross-references: GB:M94375; NID:9157461; PID:9157462
A:Note: sequence extracted from NCBI backbone (NCBIN:111933, NCBIP:111934)
C:Genetics:
A:Gene: FLYBase:Flyr2
A:Cross-references: FLYBase:FBgn0004598
C:Superfamily: subtilisin homology
C:Keywords: hydrolase; serine proteinase; transmembrane protein
F:409-652/Domain: subtilisin homology <SBT>
F:418-457,638/Active site: Asp, His, Ser #status predicted

Query Match 11.6%; Score 176.5; DB 2; Length 1680;
Best Local Similarity 28.0%; Pred. No. 0.0003;
Matches 60; Conservative 24; Mismatches 77; Indels 53; Gaps 12;

Db 3 HLRLSMFLILNFMFYGSONASRGRRMRHNPVSGCGGATGATSDY-NGCLSCRR 61
1038 HLHVID-LAVLQFCPPGYFENS-----RNRICVP-----CEPNCAACODHPEYCTSCDH 1087
Qy 62 LEFALERIGMKQIVCLSSCPGYGTRYPDINKCKKADCTCF--NKNFTCKKSGF 119
1088 LVMHEK-----CYSACPIDTYET--EDNKAFCSTCATCNGPTDODICRCSR 1136
Qy 120 YLHLCKCLDNCPEGLFANHTMECVSIHCEVSEWNPSPCTKKGKTCGFKGRTETRY-R 178
1137 YAMOKKCLISPDGFFADKKRLECM-----PQEQCKTC-----TSNGVCS 1177
Qy 179 ELIHPKSAKMLCPPTNTRKCTVQOKK-CQKGE 211
1178 ECLQMT-----LNKRDKCIVSGEGCGSESE 1203

RESULT 6
A39490
subtilisin-like proprotein convertase (EC 3.4.21.-) PACE4 precursor, splice form A - hum
N:Alternate names: kexin homolog
C:Species: Homo sapiens (man)
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 31-Mar-2000
C:Accession: A39490
R:Kiefer, M.C.; Tucker, J.E.; Joh, R.; Landsberg, K.E.; Saltman, D.; Barr, P.J.
DNA Cell Biol. 10, 757-769, 1991
A:Title: Identification of a second human subtilisin-like protease gene in the fes/fps r
A:Reference number: A39490; MUID:92075167; PMID:1741956
A:Accession: A39490
A:Molecule type: mRNA
A:Residues: 1-969 <KIE>
A:Cross-references: GB:M80482; NID:9189531; PIDN:AAA59998.1; PID:9189532
C:Genetics:
A:Gene: GDB:PACE4
A:Cross-references: GDB:131390; OMIM:167405
A:Map position: 15q26-15q28
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: alternative splicing; hydrolase; serine proteinase
F:150-969/Product: serine proteinase PACE4 #status predicted <SIG>
F:196-444/Domain: subtilisin homology <SBT>
F:203,246,420/Active site: Asp, His, Ser #status predicted

Query Match 11.4%; Score 173.5; DB 1; Length 969;
Best Local Similarity 28.8%; Pred. No. 0.00032;
Matches 44; Conservative 16; Mismatches 54; Indels 39; Gaps 7;

Db 19 YIGSONASRGRRMRHNPVSGCGGATGATSDY-NGCLSCRRLEFALERIGMKQIV 76
737 YFGDTAARRRR-----CHKGETSSRAAQCCLSC-RGTY-----HHQENMT 779
Qy 77 CLSSCPGYGTRYPDINKCKKADCTCFNK-NFTCKKSGFYHLKCLDNCPEGLE 135
780 CVTLCPAGFYADE--SQKCLCKHPSCKKCVDEPECKTVCKEGFSIARSGCIPDCEPTY 837
Qy 136 ANHTMECVSIHCEVSEWNPSPCTKKGKTCG 168

Db 838 FDSLEIRGECCH-----TCG 853

RESULT 7
B48225
probable proprotein convertase (EC 3.4.21.-) 5 precursor - rat
N:Alternate names: PCS precursor
C:Species: Rattus norvegicus (Norway rat)
C:Date: 26-May-1994 #sequence_revision 26-May-1994 #text_change 23-Feb-1997
C:Accession: B48225
R:Russon, U.; Vleau, D.; Hamelin, J.; Day, R.; Chretien, M.; Seidah, N.G.
Proc. Natl. Acad. Sci. U.S.A. 90, 6691-6695, 1993
A:Title: cDNA structure of the mouse and rat subtilisin/kexin-like PCS: a candidate pr
A:Reference number: A48225; MUID:93342056; PMID:8341687
A:Accession: B48225
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-915 <LUS>
A:Cross-references: GB:U14933
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
C:Keywords: duplication; glycoprotein; hydrolase; integrin binding; serine proteinase
F:1-34/Domain: signal sequence #status predicted <SIG>
F:35-915/Product: probable proprotein convertase 5 #status predicted <PRO>
F:117-915/Product: probable proprotein convertase 5 #status experimental <MAT>
F:164-402/Domain: subtilisin homology <SBT>
F:173,214,388/Active site: Asp, His, Ser #status predicted

Query Match 11.1%; Score 168; DB 2; Length 915;
Best Local Similarity 26.3%; Pred. No. 0.0007;
Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;

Qy 36 PNVSQ-GCGG---GCATCSDYNGCLSCRRLEFALERIGMKQIVCLSSCPGYGTRY 90
640 PECSVEGCDGPGPHDCTCLHYHKLGNTR-----ICVSCPPGHF---H 682
Db 91 PDINKCKKADCTCFNKNF--CTKCKSGFYH--LGKCLDNCPEGLFANHTMECVSI 146
683 ADKKRRKCAPKPCSCFSHADQCLSKGYFLNERTSCVAGCEGSGYQIKKNT--- 739
Qy 147 VHCVESEWNPSPCTKKGKTC-GFKGRTETRYEIIQHPKSAKMLCPPTNTRKCTVQOK 205
Db 740 -----CKSECKTCTGPHNCTE-----CKGGL---SLGSRCSV--- 771
Qy 206 KCKGE 211
Db 772 TCEDGQ 777

RESULT 8
152527
PACE4 - mouse (fragment)
C:Species: Mus musculus (house mouse)
C:Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 16-Jul-1999
C:Accession: 152527
R:Hosaka, M.; Murakami, K.; Nakayama, K.
Biomed. Res. 15, 383-390, 1994
A:Title: PACE4 is a ubiquitous endoprotease that has similar but not identical substra
A:Reference number: 152527
A:Accession: 152527
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-932 <RES>
A:Cross-references: GB:D50060; NID:9769700; PIDN:BA08777.1; PID:9769701
C:Superfamily: subtilisin-like proteinase PACE4; subtilisin homology
F:112-810/Domain: subtilisin homology <SBT>

Query Match 11.0%; Score 167.5; DB 2; Length 932;
Best Local Similarity 24.1%; Pred. No. 0.00076;
Matches 52; Conservative 25; Mismatches 78; Indels 61; Gaps 11;

Qy 19 YIGSONASRGRRMRHNPVSGCGGATGATSDY-NGCLSCRRLEFALERIGMKQIV 76

[illegible]

Db 724 THCPDSYVDTKX---NLCRCSCENCKTCTEPHNCCTEGCDLGLGSCSCSVCSDGRYFN 780
 Qy 138 NHTME-----CVS-----YHCEVSEV-----NPMSPCTKK 163
 Db 781 GDDCGCHRFCAATGAGAGADGCTEGYEMEDRCVSCSTSYFDHSSENGKSKCK 840
 Qy 164 GKTC-----GFKRGTE-----TRVEI1QHPSAKGNLCPTNETRKTCTY 202
 Db 841 DISCLTNGPGRFNCTSPSGYLLDGMQMGALCKDADTESWAGGFCMLVKKNNLC-- 898
 Qy 203 QRRKCK 209
 Db 899 QRRKVLQ 905

RESULT 12

T24232
 hypothetical protein R17.3 - Caenorhabditis elegans
 C:Species: Caenorhabditis elegans
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 21-Jan-2000
 C:Accession: T24232
 R:Barlow, K.
 submitted to the EMBL Data Library, March 1997
 A:Reference number: Z19860
 A:Accession: T24232
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-440 <Wtl>
 A:Cross-references: EMBL:Z92809; PIDN: CAB07269.1; GSPDB: GN00021; CESP: R17.3
 A:Experimental source: clone R17
 C:Genetics:
 A:Gene: CESP:R17.3
 A:Map position: 3
 A:Insertions: 36/3; 89/3; 179/2; 241/1; 306/1; 346/2
 C:Superfamily: Caenorhabditis elegans hypothetical protein R17.3

Query Match 10.6%; Score 161; DB 2; Length 440;
 Best Local Similarity 23.7%; Pred. No. 0.0012;
 Matches 53; Conservative 30; Mismatches 87; Indels 54; Gaps 8;
 Qy 59 KPRLPFALERIGMKGIVCLSSCPSGYGTGRYPDINKCTCKADCTCFNNKCTCKSG 118
 Db 172 KPRHLIRYSLSKFMPKVTSTLYENRVOFANNLYLESSISECYDEHC----- 225
 Qy 119 FYLHIGKCLDNPEGLENNHTMECVSIHCEVSEVNPSPCTKKGTCTGCTETRRV 178
 Db 226 --VTLGDC-----CSDTFVCP-RDCVLTDSDHTQCTADNGTCGL--GTQKRLK 271
 Qy 179 EIIQHPSAKGNLCPTNETRKTCTVQRKCKQKGRGKGRERKPKNGESKEALPDSKS 238
 Db 272 HVIQHAERGAACEPLKEMRTCFVE--CR-----PKKSALD--- 306
 Qy 239 LESSKEIPEORENKOOKKRR-----YDDKQSVSVSTVH 273
 Db 307 ITTVALILDYRNKTRSKIRNNIYMDLPVNAEKMKKATYYCVH 350

RESULT 13

A59180
 Mnt inhibitory factor-1 - human
 C:Species: Homo sapiens (man)
 C:Date: 18-Feb-2000 #sequence_revision 18-Feb-2000 #text_change 18-Feb-2000
 C:Accession: A59180
 R:Heien, J.C.; Kodjabachian, L.; Rebber, M.L.; Ratner, A.; Smallwood, P.M.; Samos, C.H.
 Nature 398, 431-438, 1999
 A:Title: A new secreted protein that binds to Wnt proteins and inhibits their activities
 A:Reference number: A59180; MUID: 99215557; PMID: 10201374
 A:Accession: A59180
 A:Status: preliminary; not compared with conceptual translation
 A:Molecule type: mRNA
 A:Residues: 1-379 <HST>
 A:Cross-references: GB: A122922; NID: g4585369; PIDN: AAD25402.1; PID: g4585370

Query Match 10.5%; Score 158.5; DB 2; Length 379;
 Best Local Similarity 24.9%; Pred. No. 0.0015;
 Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Qy 42 CQGGCAT--CSDYNC-----LSCKPRLPFALERIGMKGIVCLSS---C 81
 Db 182 CPGCGRNGGFCNBRICECPDGFHGHCEKALCTPRONG-----GLCVTFPGFIC 232
 Qy 82 PSGYGTGRYPDINKCTKC-KADCD--TCFNKRFCTKCKSGFYHLGKCLDNPEGLEANNH 139
 Db 233 PGRFG-----VNCIKANCSTTCNNGTC-----FY--PGKCI--CPGLEGE-- 271
 Qy 140 TMECVSIHCEVSEVNPSPCTKGTGFRGTETRVREI1QHPSAKGNLCPT----- 193
 Db 272 -----OCEISKCP--QPCRMGGKCI G--KSKCKSKGYQGLCSKPCVCEPGCGAHG 318
 Qy 194 --TNETRKTCTVQRKCKQKGRGKGRERKX-----KPNKESKEALPDSKLESSKEIP 246
 Db 319 TCHERNK-----QCEGHHGHCHMKRYEASLIHALRPAQAOIRQHTPLSKAEERRRPP 373
 Qy 247 E 247
 Db 374 E 374

RESULT 14

153282
 gene PACB4 protein - rat
 C:Species: Rattus norvegicus (Norway rat)
 C:Date: 02-Aug-1996 #sequence_revision 02-Aug-1996 #text_change 22-Jun-1999
 C:Accession: 153282
 R:Johnson, R.C.; Darlington, D.N.; Hand, T.A.; Bloomquist, B.T.; Mains, R.E.
 Endocrinology 135, 1178-1185, 1994
 A:Title: PACB4: a subtilisin-like endoprotease prevalent in the anterior pituitary and
 A:Reference number: 153282; MUID: 99349873; PMID: 8070361
 A:Accession: 153282
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: mRNA
 A:Residues: 1-937 <RES>
 A:Cross-references: GB: L31894; NID: g496221; PIDN: AAA61987.1; PID: g496222
 C:Genetics:
 A:Gene: PACB4
 C:Superfamily: subtilisin-like proteinase PACB4; subtilisin homology
 F.177-415/Domain: subtilisin homology <SBT>

Query Match 10.3%; Score 156.5; DB 2; Length 937;
 Best Local Similarity 23.6%; Pred. No. 0.004;
 Matches 51; Conservative 24; Mismatches 80; Indels 61; Gaps 11;
 Qy 19 YIGSQNASRGRORRMHNVSSQCGGATCSDYN--GCLSKPRLPFALERIGMKGIV 76
 Db 705 YFGDTAARCR-----CHKGETCTGRSPQCLSCR-RGFY-----HHQETNT 747
 Qy 77 CLSSCPSGYGTYPDINKCTCKADCTCFNK--NFTCKSGFYHLGKCLDNPEGLE 135
 Db 748 CVTLCPAGLVADESQRL--CLKCHPSCKQCVDEPKSTVCKESFLARSCCLPDEPETY 805
 Qy 136 ANNTMEC-----VSIVHCEVS---EWNPMSPCTKKGTCTGCFKGTETR 176
 Db 806 FDSRLIRGCECHHTCTCTGSPRECIHCAKSFHFQDMKVPACGE-----GF----- 853
 Qy 177 VREI1QHPSAKGNLCPTNETRKTCTVQRKCKQKGR 212
 Db 854 -----YPEMPGL--PKVCRRCDCNCLSCGSSR 881

RESULT 15

A47723
 F-spondin precursor - African clawed frog
 C:Species: Xenopus laevis (African clawed frog)
 C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 17-Nov-2000
 C:Accession: A47723

R;Ruiz i Altaba, A.; Cox, C.; Jessell, T.M.; Klar, A.

Proc. Natl. Acad. Sci. U.S.A. 90, 8268-8272, 1993

A;Title: Ectopic neural expression of a floor plate marker in frog embryos injected with

A;Reference number: A47723; MUID:93376785; PMID:8367492

A;Accession: A47723

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-803 <RU1>

A;Cross-References: G6:L09123; NID:9409244; PIDN:AAA19105.1; PID:9409245

C;Superfamily: F-spondin; thrombospondin type 1 repeat homology

F;435-489/Domain: thrombospondin type 1 repeat homology <THR2>

F;607-662/Domain: thrombospondin type 1 repeat homology <THR1>

Query Match

Best Local Similarity 10.3%; Score 156; DB 2; Length 803;

Matches 58; Conservative 35; Mismatches 91; Indels 52; Gaps 12;

```

Oy 38 VSOGCGGCGATCSDYNGCLSCPRLPFALEIRIQMKGIVGCLSSCP--GYGTRYPDINK 95
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 544 VNEECPPSSCIIVTEWAEWEECS-----ATCRMGKKRRHRIMKTPADGSMCKADTTEVEK 598
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 96 C--TKCK-----ADCDTCFNQNPCTKCKSGFYHLGKCLDNCPEGLEANN--- 138
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 599 CMWPECHTIPCVLSPMSFMSDCSVTCCKGTRTRQR---MLKSPSELGDCNELELKQVEK 655
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 139 -HMECVSIWCEVSEANPWSPECTKKGKTGCFKGTETRVREIIQHPSAKGNLCPTINET 197
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 656 CMLPECP--ISCLEWYSWSEC---NKSCG--KGHMRTRMTWEPQFGAVCPETVQR 708
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Oy 198 RKCTVORKKCKGKRGKRRKKRKKPKNGESKEAIPDSKSLSSKEIPEQRENKQ 253
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 709 KRCRL--RCKCK---SSGNRRHLK-----DAREKRRSEKIKEDSDGEQ 747
   |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

Search completed: May 6, 2003, 14:53:03

Job time : 20 secs

GenCore version 5.1.4_p5_4578
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2003, 14:49:17 ; Search time 37 Seconds
(without alignments)
983.173 Million cell updates/sec

Title: US-09-894-912A-13
 Perfect score: 1516
 Sequence: 1 MGHRLRLISWLFILNFMET.....QQKKRKVQDQKQSVSVTH 273

Scoring table:	BLOSUM62	Ganor 10.0	Ganext 0.5
----------------	----------	------------	------------

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

```
Post-processing: Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries
```

1. A_Geneseq.101002.*
2. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1980.DAT.*
3. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1981.DAT.*
4. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1982.DAT.*
5. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1983.DAT.*
6. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1984.DAT.*
7. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1986.DAT.*
8. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1987.DAT.*
9. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1988.DAT.*
10. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1989.DAT.*
11. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1990.DAT.*
12. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1991.DAT.*
13. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1992.DAT.*
14. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1993.DAT.*
15. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1994.DAT.*
16. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1995.DAT.*
17. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1996.DAT.*
18. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1997.DAT.*
19. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1998.DAT.*
20. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA1999.DAT.*
21. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA2000.DAT.*
22. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA2001.DAT.*
23. /SID2/gcgcatat/geneseq/geneseqp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	length	DB	ID	Description
1	15.6	100.0	273	22	AAE13151	Human stem cell g
2	15.05	99.3	272	22	AAE13150	Human stem cell g
3	15.05	99.3	272	22	AAE13168	Human stem cell g
4	15.05	99.3	272	22	AAW78228	Human protein seq
5	15.05	99.3	272	22	AAH92220	Human protein
6	14.2	97.1	265	22	AAE13163	Human thymospond
7	14.2	97.1	292	20	AAH85607	Human secreted prot
8	14.2	97.1	292	22	AAE13170	Secreted protein c
9	14.2	97.1	292	23	ABH61846	Human SCR-1 relat
10	14.2	95.8	292	22	AAH93875	Human polypeptide
						Human protein sequ

11	1400	92.3	254	22	AAE313173	Human mature stem
12	1310.5	86.4	279	22	AAE313157	Mouse stem cell gsm
13	903	59.6	160	22	AAE313149	Human stem cell gsm
14	903	59.6	160	22	AAE313174	Human secreted pte
15	903	59.6	160	22	AAE793132	Human protein SEQ
16	638	42.1	229	22	AAE313165	Mouse chromospon
17	276	18.2	46	22	AAE313155	Human SCR-1 prote
18	221	14.1	37	22	AAE313159	Human SCR-1 prote
19	126	11.6	1679	22	AAE313157	Human SCR-1 prote
20	176.5	11.6	1679	22	ABBB60548	Drosophila melano
21	176.5	11.6	1679	22	ABBB60548	Drosophila melano
22	173.5	11.4	969	14	AAE13538	Paired basic amino
23	173.5	11.4	969	23	AAE13538	Human subcillin J
24	162.5	10.7	132	20	AAE08453	Rat serine proteas
25	162	10.7	799	20	AAE18073	CD4+ T-lymphocyte
26	162	10.7	799	20	AAE18073	CD4+ T-lymphocyte
27	162	10.7	799	20	AAE18073	CD4+ T-lymphocyte
28	162	10.7	881	20	AAE739932	CD4+ T-lymphocyte
29	162	10.7	935	19	AAE52306	T-lymphocyte prote
30	162	10.7	935	19	AAE18071	Human pro-protein
31	162	10.7	935	20	AAE18071	CD4+ T-lymphocyte
32	161	10.6	935	20	AAE18071	T-lymphocyte prote
33	160.5	10.6	288	19	AAE34557	hTcP protein fragm
34	160.5	10.6	216	20	AAE084432	Rat serine protease
35	159	10.5	479	22	AAU16960	Human novel secret
36	158.5	10.5	193	20	AAE084431	Rat serine protease
37	158.5	10.5	379	20	AAE08665	Human EGF-like hom
38	158.5	10.5	379	20	AAE13345	Human EGF-like hom
39	158.5	10.5	379	21	AAE13345	Human EGF-like hom
40	158.5	10.5	379	21	AAE13345	Human EGF-like hom
41	158.5	10.5	379	21	AAE13345	Human EGF-like hom
42	158.5	10.5	379	21	AAE13345	Human EGF-like hom
43	158.5	10.5	379	21	AAE13345	Human EGF-like hom
44	158.5	10.5	379	21	AAE13345	Human EGF-like hom
45	158.5	10.5	379	22	AAE13345	Human EGF-like hom

ALIGNMENTS

RESULT 1
AAE13151
ID AAE13151 standard; Protein; 273 AA

DT 28-JAN-2002 (first entry)

DE Human stem cell growth factor-like protein #3.

Human; sem cell growth factor-like protein; antiinflammatory; neutropenic
neutrophilic; leukopenia; immunodeficiency; immunosuppressive;
neuroprotective; vulnerrary; cytostatic; anticonvulsant; immunostimulant;
vasorelaxant; vasodilator; demyelinating; tranquilizer; cerebroprotective;
neuropathic; immunodeficiency syndrome; chronic granulomatous disease;
osteopathic; immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
duplicated immunodeficiency syndrome; acquired aplastic anemia; thalassemia;
acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
arterial white matter degeneration; anaemia; neurodegenerative disease;
Parkinson's disease; Alzheimer's disease; thrombocytopenia; SCID;
multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
autoimmune pulmonary inflammation; cancer; tumour; gene therapy;
GCR-1; supporting factor for the proliferation of stem cell.

OS	Homo sapiens.
XX	
PN	MO200177169-A2.
XX	
XX	18-OCT-2001.
PD	
PF	05-APR-2001; 2001MO-US11208
XX	
PR	05-APR-2000; 2000US-0543774

PR 28-JUN-2000; 2000US-215733P.
 PR 09-JAN-2001; 2001US-075756Z.
 PR 05-FEB-2001; 2001US-266614P.
 XX
 PA (HYSE-) HYSEQ INC.
 PA (KIRI) KIRIN BEER KK.
 XX
 PI Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Dmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX
 DR WPI, 2001-657166/75.
 DR N-PSDB; AAD21725.
 XX
 PT Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis -
 XX
 PS Claim 28; Page 214-215; 232pp; English.
 XX
 CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell
 CC growth factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety
 CC of diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
 CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
 CC storage diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the
 CC invention are useful for treating diseases such as Parkinson's
 CC disease, Alzheimer's disease and other neurodegenerative diseases,
 CC chromocytopenia, immune deficiencies and disorders such as severe
 CC combined immunodeficiency (SCID) and autoimmune disorders such as
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC and autoimmune pulmonary inflammation. Sequences of the invention are
 CC also useful in gene therapy. The present sequence is stem cell growth
 CC factor-like protein from human.
 CC
 XX
 SO Sequence 273 AA:
 Query Match 100.0%; Score 1516; DB 22; Length 273;
 Best Local Similarity 100.0%; Pred. No. 3e-112;
 Matches 273; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MGLRLISLWFIILFMFETISQNASGRORRMHNVSOQGGCATCSYNGCLSCRP 60
 DB 1 MGLRLISLWFIILFMFETISQNASGRORRMHNVSOQGGCATCSYNGCLSCRP 60
 QY 61 RLFFALERIGMKGIVCLSSCPSGYGYTRYPDINKTKKADCDCTFNKRNCTKCKSGFY 120
 DB 61 RLFFALERIGMKGIVCLSSCPSGYGYTRYPDINKTKKADCDCTFNKRNCTKCKSGFY 120
 QY 121 LHLGKLDLCPGEGEANNHTMECVSIHCEVSEPMWSPCTKKGKTCGFKGTETRREI 180
 DB 121 LHLGKLDLCPGEGEANNHTMECVSIHCEVSEPMWSPCTKKGKTCGFKGTETRREI 180
 QY 181 IOHPSAKKGLCPPTNETKCTVOKKKOKERKKRERKKRPPKSGSKAIPDSKSL 240
 DB 181 IOHPSAKKGLCPPTNETKCTVOKKKOKERKKRERKKRPPKSGSKAIPDSKSL 240
 QY 241 SSKEIPEORENKKOQKRRKRVODKXSVSVTVH 273
 DB 241 SSKEIPEORENKKOQKRRKRVODKXSVSVTVH 273
 RFSULT 2

AAE13150
 ID AAE13150 standard; Protein: 272 AA.
 AC AAE13150;
 XX
 DT 28-JAN-2002 (first entry)
 XX
 DE Human stem cell growth factor-like protein #2.
 KW Human; stem cell growth factor-like protein; antiinflammatory; nocotropic;
 KW neuroprotective; vulnerary; cytosolic; anticonvulsant; immunostimulant;
 KW vasotropic; virucide; dermatological; tranquilliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy;
 KW SCR-1; supporting factor for the proliferation of stem cell.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Peptide 1..21
 FT Protein /label= Signal_peptide
 FT 22..272 /note= "Human mature stem cell growth
 FT factor-like protein"
 PN MO200177169-A2.
 PD 18-OCT-2001.
 XX
 PF 05-APR-2001; 2001WO-US11206.
 XX
 PR 05-APR-2000; 2000US-0543774.
 PR 28-JUN-2000; 2000US-215733P.
 PR 09-JAN-2001; 2001US-075756Z.
 PR 05-FEB-2001; 2001US-266614P.
 XX
 PA (HYSE-) HYSEQ INC.
 PA (KIRI) KIRIN BEER KK.
 XX
 PI Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Dmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX
 DR WPI, 2001-657166/75.
 DR N-PSDB; AAD21724.
 XX
 PT Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis -
 XX
 PS Claim 28; Page 211-212; 232pp; English.
 XX
 CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell
 CC growth factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety
 CC of diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
 CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal

CC storage diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the
 CC invention are useful for treating diseases such as Parkinson's
 CC disease, Alzheimer's disease and other neurodegenerative diseases,
 CC chromocytopenia, immune deficiencies and disorders such as severe
 CC combined immunodeficiency (SCID) and autoimmune disorders such as
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC and autoimmune pulmonary inflammation. Sequences of the invention are
 CC also useful in gene therapy. The present sequence is stem cell growth
 CC factor-like protein from human.

XX
 XX Sequence 272 AA:

Query Match 99.3%; Score 1505; DB 22; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2,3e-111;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATGSDVNGCLSCPKPL 62
 DB 2 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATGSDVNGCLSCPKPL 61

QY 63 FFALEIRIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCFKCKSGFYHL 122
 DB 62 FFALEIRIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCFKCKSGFYHL 121

QY 123 LGKCLDNCEGLLEANNHTMECVSIHCEVSEWNPSPCTKKGKTCGFRGTERVREIIIO 182
 DB 122 LGKCLDNCEGLLEANNHTMECVSIHCEVSEWNPSPCTKKGKTCGFRGTERVREIIIO 181

QY 183 HPSAKGNLCPTNETRKTCTVQRKKCKGGRKKGRERKPKNGSKESKEAIPDSKLSLESS 242
 DB 182 HPSAKGNLCPTNETRKTCTVQRKKCKGGRKKGRERKPKNGSKESKEAIPDSKLSLESS 241

QY 243 KEIPEORENKQOQKKRVQDKQKSVSVSTVH 273
 DB 242 KEIPEORENKQOQKKRVQDKQKSVSVSTVH 272

RESULT 3
 ID AAE13168
 XX AAE13168 standard; Protein; 272 AA.
 AC AAE13168;
 XX
 DT 28-JAN-2002 (first entry)
 XX
 DE Human stem cell growth factor-like protein #4.
 XX
 KW Human, stem cell growth factor-like protein; antiinflammatory; noctropic;
 KW neuroprotective; vulnerary; cyostatic; anticonvulsant; immunostimulant;
 KW vasotrophic; virucide; dermatological; tranquilliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Miskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy;
 KW SCR-1; supporting factor for the proliferation of stem cell.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Signal_peptide
 FT Protein 22..272 "Human mature stem cell growth
 FT /note="Human mature stem cell growth
 XX factor-like protein"
 XX
 PN W0200177169-A2.

PD 18-OCT-2001.
 XX
 PF 05-APR-2001; 2001WO-US11208.
 XX
 PR 05-APR-2000; 2000US-0543774.
 PR 28-JUN-2000; 2000US-215733P.
 PR 09-JAN-2001; 2001US-0757562.
 PR 05-FEB-2001; 2001US-266614P.
 PA (HYSE-) HYSEO INC.
 PA (KIRI) KIRIN BEER KK.
 XX
 PI Tang TY, Labat I, Tillinghaast JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Crain B, Dickson W, Mize NK, Nishikawa M;
 DR N-PSDB; AAD21728.
 DR WPI; 2001-657166/75.
 PT Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis
 XX
 XX Claim 28; Page 226-227; 232pp; English.

CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably, primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell
 CC growth factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety
 CC of diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Miskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
 CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
 CC storage diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the
 CC invention are useful for treating diseases such as Parkinson's
 CC disease, Alzheimer's disease and other neurodegenerative diseases,
 CC chromocytopenia, immune deficiencies and disorders such as severe
 CC combined immunodeficiency (SCID) and autoimmune disorders such as
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC and autoimmune pulmonary inflammation. Sequences of the invention are
 CC also useful in gene therapy. The present sequence is stem cell growth
 CC factor-like protein from human.

XX
 XX Sequence 272 AA:

Query Match 99.3%; Score 1505; DB 22; Length 272;
 Best Local Similarity 100.0%; Pred. No. 2,3e-111;
 Matches 271; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 3 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATGSDVNGCLSCPKPL 62
 DB 2 HRLISWLFILINMEYIGSONASRGRRORRMHPNVSGCCGGCATGSDVNGCLSCPKPL 61

QY 63 FFALEIRIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCFKCKSGFYHL 122
 DB 62 FFALEIRIGMKQIGVCLSSCPGYYGTRYPDINKCTCKACDCTCFNNKFCFKCKSGFYHL 121

QY 123 LGKCLDNCEGLLEANNHTMECVSIHCEVSEWNPSPCTKKGKTCGFRGTERVREIIIO 182
 DB 122 LGKCLDNCEGLLEANNHTMECVSIHCEVSEWNPSPCTKKGKTCGFRGTERVREIIIO 181

QY 183 HPSAKGNLCPTNETRKTCTVQRKKCKGGRKKGRERKPKNGSKESKEAIPDSKLSLESS 242
 DB 182 HPSAKGNLCPTNETRKTCTVQRKKCKGGRKKGRERKPKNGSKESKEAIPDSKLSLESS 241

QY 243 KEIPEORENKQOQKKRVQDKQKSVSVSTVH 273

DB 242 KEIPEORENKQKKRKVKQDKQKSVSVTVH 272

RESULT 4
ID AAM78328 standard; Protein; 272 AA.
XX
XX AAM78328;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human protein SEQ ID NO 990.
XX
XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
KW vaccine; peptide therapy; stem cell growth factor; haematopoiesis;
KW tissue growth factor; immunomodulatory; cancer; leukaemia;
KW tissue system disorder; arthritis; inflammation.
XX
OS Homo sapiens.
XX
PN WO200157190-A2.
XX
PD 09-AUG-2001.
XX
XX 05-FEB-2001; 2001WO-US04098.
XX
PF 03-FEB-2000; 2000US-0496914.
PR 27-APR-2000; 2000US-0560875.
PR 20-JUN-2000; 2000US-0598075.
PR 19-JUL-2000; 2000US-0620325.
PR 01-SEP-2000; 2000US-0649356.
PR 15-SEP-2000; 2000US-0663561.
PR 20-OCT-2000; 2000US-0693325.
PR 30-NOV-2000; 2000US-0728422.
XX
XX (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y;
PI Zhao QA, Wang D, Zhang J, Ren F, Chen R, Wang ZM;
PI Xue AJ, Yang Y, Wehrman T, Goodrich R;
XX
XX WPI; 2001-476283/51.
DR N-PSDB; AAK51461.
XX
XX Nucleic acids encoding polypeptides with cytokine-like activities,
PT useful in diagnosis and gene therapy -
XX
PS Claim 20; Page 3214-3215; 6221pp; English.
XX
XX The invention relates to polynucleotides (AAK51456-AAK51435) and the
CC encoded polypeptides (AAM78323-AAM80302) that exhibit activity elating to
CC cytokine, cell proliferation or cell differentiation or which may induce
CC production of other cytokines in other cell populations. The
CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
CC peptide therapy. The polypeptides have various cytokine-like activities,
CC e.g. stem cell growth factor activity, haematopoiesis regulating
CC activity, tissue growth factor activity, immunomodulatory activity and
CC activity/inhibin activity and may be useful in the diagnosis and/or
CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
CC inflammation.
CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
CC (AAM80020) are omitted as the relevant pages from the sequence listing
CC were missing at the time of publication.
XX
XX Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 22; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.3e-111; Indels 0; Gaps 0;
Matches 271; Conservative 0; Mismatches 0;

3 HLRILSWLFIILNFMFYIGSONASGRGRORRMHPNVSGCGGCAATCSDYNGCLSCPKRL 62

DB 2 HLRILSWLFIILNFMFYIGSONASGRGRORRMHPNVSGCGGCAATCSDYNGCLSCPKRL 61

QY 63 FEALERTGMKQIGVCLSSCPGSGYGYTRYPDINKCTKQKADCDPTCFNKPCTKCKSGFYLH 122

DB 62 FFALERIGMKQIGVCLSSCPGSGYGYTRYPDINKCTKQKADCDPTCFNKPCTKCKSGFYLH 121

QY 123 LGKCLDNCPEGLEANNHTMECVSIVHCEVSEBNPMSPTCKGKTGFRGTETRVREIIQ 182

DB 122 LGKCLDNCPEGLEANNHTMECVSIVHCEVSEBNPMSPTCKGKTGFRGTETRVREIIQ 181

QY 183 HPSAKGNLCPTNETPRTKCTVORKKCKGGRKGRERKRRKPKNGESKEAIPDSKLESS 242

DB 182 HPSAKGNLCPTNETPRTKCTVORKKCKGGRKGRERKRRKPKNGESKEAIPDSKLESS 241

QY 243 KEIPEORENKQKKRKVKQDKQKSVSVTVH 273

DB 242 KEIPEORENKQKKRKVKQDKQKSVSVTVH 272

RESULT 5
ID AAB99220 standard; Protein; 272 AA.
XX
XX AAB99220;
XX
XX 07-SEP-2001 (first entry)
XX
DE Human thrombospondin-30.
XX
KW Human; thrombospondin-30; cytostatic; anti-HIV; antiinflammatory;
KW malignant tumour; haemopathy; HIV infection; immunological disease;
KW inflammation disease.
XX
OS Homo sapiens.
XX
PN WO200140294-A1.
XX
PD 07-JUN-2001.
XX
XX 27-NOV-2000; 2000WO-CN00494.
XX
XX 29-NOV-1999; 99CN-0124148.
XX
XX (BIOR-) BIORAD GENE DEV LTD SHANGHAI.
XX
PI Mao Y, Xie Y;
XX
XX WPI; 2001-397948/42.
DR N-PSDB; AAB45131.
XX
XX Human thrombospondin-30 and polynucleotide is useful in diagnosis and
PT treatment of, e.g., malignant tumor, hemopathy, HIV infection,
PT immunological diseases and various inflammatory diseases -
XX
PS Claim 1; Page 27-28; 33pp; Chinese.
XX
XX The present sequence is the protein sequence for human thrombospondin-30.
CC Thrombospondin-30 protein and coding sequence are useful in the diagnosis
CC and treatment of malignant tumour, haemopathy, HIV infection,
CC immunological diseases and various inflammatory diseases. In addition
CC thrombospondin-30 protein may be used for screening mutants, agonists,
CC antagonists or inhibitors, or for use in peptide fingerprinting
CC identification. The thrombospondin-30 coding sequence may be used as
CC primers for nucleic acid amplification reaction or as probes for
CC hybridisation reaction, or in producing gene chips or microarrays.
XX
XX Sequence 272 AA;

Query Match 99.3%; Score 1505; DB 22; Length 272;
Best Local Similarity 100.0%; Pred. No. 2.3e-111; Indels 0; Gaps 0;
Matches 271; Conservative 0; Mismatches 0;

3 HLRILSWLFIILNFMFYIGSONASGRGRORRMHPNVSGCGGCAATCSDYNGCLSCPKRL 62

```

|||||
Db 2 HRLRLSWLFIILNFMETIGSONASRGRRMRHNPVSGCCGGCATCSDDYNGCLSCRPRL 61
Oy 63 PALERIRKMGKQIGVCLSSCPGSGYGTGRYPDINKCTKCAADCDTCFNKPFCTKSGPYLH 122
Db 62 PALERIRKMGKQIGVCLSSCPGSGYGTGRYPDINKCTKCAADCDTCFNKPFCTKSGPYLH 121
Oy 123 LGKCLDNCPEGLLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGKRGTEFRVREIIQ 182
Db 122 LGKCLDNCPEGLLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGKRGTEFRVREIIQ 181
Oy 183 HPSAKGNLCPTNTRKCTVORRKCCKGGRKKGERKKRKNPKNGESKEAIPDSKLESS 242
Db 182 HPSAKGNLCPTNTRKCTVORRKCCKGGRKKGERKKRKNPKNGESKEAIPDSKLESS 241
Oy 243 KEIPEORENKQOQKRRKVDKOKSVSVSTVH 273
Db 242 KEIPEORENKQOQKRRKVDKOKSVSVSTVH 272

RESULT 6
AAE13163
ID AAE13163 standard; Protein; 265 AA.
AC AAE13163;
XX
XX
DT 28-JAN-2002 (first entry)
DE Human secreted protein from clone DA228_6.
XX
XX
Human; stem cell growth factor-like protein; antiinflammatory; neutrotrophic;
neuroprotective; vulnerary; cytoskeletal; anticonvulsant; immunostimulant;
vasotrophic; virucide; dermatological; tranquiliser; cerebroprotective;
osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
duplicated immunodeficiency syndrome; Miskot-Aldrich syndrome; AIDS;
acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
adrenal white matter degeneration; anaemia; neurodegenerative disease;
Parkinson's disease; Alzheimer's disease; thrombocytopenia; SCID;
severe combined immunodeficiency; immune disorder; autoimmune disease;
multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1;
supporting factor for the proliferation of stem cell; secreted protein.
XX
XX Homo sapiens.
OS
XX
XX WO200177169-A2.
XX
XX 18-OCT-2001.
XX
XX
XX 05-APR-2001; 2001WO-US11208.
XX
XX
XX 05-APR-2000; 2000US-0543774.
XX
XX 28-JUN-2000; 2000US-215733P.
XX
XX 09-JAN-2001; 2001US-0757562.
XX
XX 05-FEB-2001; 2001US-266614P.
XX
XX
XX (HYSE-) HYSEQ INC.
XX
XX (KIRI) KIRIN BEER KK.
XX
XX
XX Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
XX
XX Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
XX
XX WPI: 2001-657166/75.
XX
XX
XX Novel stem cell growth factor like polypeptides and polynucleotides for
XX
XX identifying modulators useful for treating diseases such as Alzheimer's
XX
XX disease, cancer, rheumatoid arthritis, osteoporosis
XX
XX
XX Disclosure: Fig 3; 23pp; English.
XX
XX
XX The patent discloses novel stem cell growth factor-like proteins and
XX
XX polynucleotides encoding them. Proteins of the invention are also known

```

```

CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
CC cell growth factor-like proteins are useful for supporting proliferation
CC or survival of a stem cell or germ cell which is preferably primordial
CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
CC cell. The haematopoietic progenitor cell cultured using stem cell
CC growth factor-like proteins can replace as a graft for the bone marrow
CC transplantation or cord blood transplantation for treating a variety
CC of diseases such as immunodeficiency syndrome, chronic granulomatous
CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
CC Miskot-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
CC storage diseases such as mucopolysaccharidosis, adrenal white matter
CC degeneration, a variety of cancer and tumours. Proteins of the
CC invention are useful for treating diseases such as Parkinson's
CC disease, Alzheimer's disease and other neurodegenerative diseases,
CC thrombocytopenia, immune deficiencies and disorders such as severe
CC combined immunodeficiency (SCID) and autoimmune disorders such as
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC and autoimmune pulmonary inflammation. Sequences of the invention are
CC also useful in gene therapy. The present sequence is human secreted
CC protein from clone DA228_6.
XX
XX
SQ Sequence 265 AA;
XX
XX
XX Query Match 97.1%; Score 1472; DB 22; Length 265;
XX
XX Best Local Similarity 100.0%; Pred. No. 9e-109;
XX
XX Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX
Oy 3 HRLRLSWLFIILNFMETIGSONASRGRRMRHNPVSGCCGGCATCSDDYNGCLSCRPRL 62
Db 2 HRLRLSWLFIILNFMETIGSONASRGRRMRHNPVSGCCGGCATCSDDYNGCLSCRPRL 61
Oy 63 PALERIRKMGKQIGVCLSSCPGSGYGTGRYPDINKCTKCAADCDTCFNKPFCTKSGPYLH 122
Db 62 PALERIRKMGKQIGVCLSSCPGSGYGTGRYPDINKCTKCAADCDTCFNKPFCTKSGPYLH 121
Oy 123 LGKCLDNCPEGLLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGKRGTEFRVREIIQ 182
Db 122 LGKCLDNCPEGLLEANNHTMECVSIHVCEVSEWNPSPCTKGGKTCGKRGTEFRVREIIQ 181
Oy 183 HPSAKGNLCPTNTRKCTVORRKCCKGGRKKGERKKRKNPKNGESKEAIPDSKLESS 242
Db 182 HPSAKGNLCPTNTRKCTVORRKCCKGGRKKGERKKRKNPKNGESKEAIPDSKLESS 241
Oy 243 KEIPEORENKQOQKRRKVDKOKSVSVSTVH 273
Db 242 KEIPEORENKQOQKRRKVDKOKSVSVSTVH 272

RESULT 7
AAW85607
ID AAW85607 standard; Protein; 292 AA.
XX
XX
XX AAW85607;
XX
XX
XX 02-MAR-1999 (first entry)
XX
XX
XX Secreted protein clone da228_6.
XX
XX
XX Clone: secreted protein; protein factor; cytokine; lymphokine;
XX
XX interferon; colony stimulating factor; CSF; interleukin; cloning;
XX
XX tumour invasion; tumour suppression; immune boosting.
XX
XX
XX Homo sapiens.
XX
XX
XX WO9849302-A1.
XX
XX
XX 05-NOV-1998.
XX
XX
XX 24-APR-1998; 98WO-US08336.
XX
XX

```

PR 23-APR-1998; 98US-0065125.
PR 25-APR-1997; 97US-0845296.
XX
PA (GENY) GENETICS INST INC.
XX
PI Agostino MJ, Jacobs K, Lavallie ER, McCoy JM, Merberg D;
PI Racie LA, Spaulding V, Treacy M;
DR WPI; 1999-024059/02.
XX N-PSDB; AAV83133.
XX
PT New polynucleotides encoding secreted human proteins - are derived
PT from human foetal brain, adult brain, adult blood or placenta cDNA
PT libraries, useful, e.g. as potential immunomodulators
XX
PS Claim 8; Page 63-64; 104pp; English.
XX
CC The nucleotide sequence (NS) of the full-length protein-coding
CC sequence of clones c1254 (AAV83132), da2286 (AAV83133), du4105
CC (AAV83134), eh801 (AAV83135), et3691 (AAV83136), fh1235 (AAV83137), fm601
CC (AAV83138) or fr4732 (AAV83139), (all clones are deposited as ATCC 98415)
CC and the proteins they encode are predicted to have biological activities
CC which would make them suitable for treating, preventing or
CC ameliorating medical conditions in humans and animals for example,
CC tumour suppression/invasion activity, immune system boosting
CC activity. The polynucleotides are also believed to be useful for
CC gene therapy.
XX
SQ Sequence 292 AA;
XX
Query Match 97.1%; Score 1472; DB 20; Length 292;
Best Local Similarity 100.0%; Pred. No. 1e-108;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 3 HLRILSWLFIILNMEYIGSONASRGRRORRHPNVSGCGCATGSDYNGLSCKRPL 62
DB 2 HLRILSWLFIILNMEYIGSONASRGRRORRHPNVSGCGCATGSDYNGLSCKRPL 61
QY 63 FFALRIGMKOIGVCLSSCPGSGYGRYPDINKCTKACADCTCFNNKFCFKKSGFYH 122
DB 62 FFALRIGMKOIGVCLSSCPGSGYGRYPDINKCTKACADCTCFNNKFCFKKSGFYH 121
QY 123 LKCLDNCPEGLEANNHTMECVSIHCEVSENNPSPCTKKGKTCGRGTERVREII 182
DB 122 LKCLDNCPEGLEANNHTMECVSIHCEVSENNPSPCTKKGKTCGRGTERVREII 181
QY 183 HPSAKGNLCPTNETRKTCTVORKKCKGGERGKGRERKPKNGESKEAIPDSKLESS 242
DB 182 HPSAKGNLCPTNETRKTCTVORKKCKGGERGKGRERKPKNGESKEAIPDSKLESS 241
QY 243 KEIPRENNKQOKKRYVODKXS 266
DB 242 KEIPRENNKQOKKRYVODKXS 265
XX
RESULT 8
AAE13170 ID AAE13170 standard; Protein; 292 AA.
XX
XX AAE13170;
AC
DT 28-JAN-2002 (first entry)
XX
XX Human SCR-1 related protein.
DE
XX Human; stem cell growth factor-like protein; antiinflammatory; neutrophic;
KM neuroprotective; vulnerary; cyostatic; anticonvulsant; immunostimulant;
KM vasotrophic; virucide; dermatological; tranquilliser; cerebroprotective;
KM osteoplastic; immunodeficiency syndrome; chronic granulomatous disease;
KM duplicated immunodeficiency syndrome; Miskot-Aldrich syndrome; AIDS;
KM acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
KM Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
KM adrenal white matter degeneration; anaemia; neurodegenerative disease;

KM Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
KM severe combined immunodeficiency; immune disorder; autoimmune disorder;
KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
KM autoimmune pulmonary inflammation; cancer; tumour; gene therapy;
KM SCR-1; supporting factor for the proliferation of stem cell.
OS unidentified.
XX
XX WO200177169-A2.
XX
XX 18-OCT-2001.
XX
XX 05-APR-2001; 2001WO-US11208.
XX
XX 05-APR-2000; 2000US-0543774.
XX 28-JUN-2000; 2000US-215733P.
PR 09-JAN-2001; 2001US-0757562.
PR 05-FEB-2001; 2001US-266614P.
XX
XX (HYSE-) HYSEQ INC.
XX (KIRI) KIRIN BEER KK.
PA
PA Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
XX WPI; 2001-657166/75.
DR N-PSDB; AAD21740.
XX
XX Novel stem cell growth factor like polypeptides and polynucleotides for
PT identifying modulators useful for treating diseases such as Alzheimer's
PT disease, cancer, rheumatoid arthritis, osteoporosis
XX
XX Claim 27; Page 231-232; 232pp; English.
XX
CC The patent discloses novel stem cell growth factor-like proteins and
CC polynucleotides encoding them. Proteins of the invention are also known
CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
CC cell growth factor-like proteins are useful for supporting proliferation
CC or survival of a stem cell or germ cell which is preferably primordial
CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
CC cell. The haematopoietic progenitor cell cultured using stem cell
CC growth factor-like proteins can replace as a graft for the bone marrow
CC transplantation or cord blood transplantation for treating a variety
CC of diseases such as immunodeficiency syndrome, chronic granulomatous
CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
CC Miskot-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
CC storage diseases such as mucopolysaccharidosis, adrenal white matter
CC degeneration, a variety of cancer and tumours. Proteins of the
CC invention are useful for treating diseases such as Parkinson's
CC disease, Alzheimer's disease and other neurodegenerative diseases,
CC thrombocytopaenia, immune deficiencies and disorders such as severe
CC combined immunodeficiency (SCID) and autoimmune disorders such as
CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
CC and autoimmune pulmonary inflammation. Sequences of the invention are
CC also useful in gene therapy. The present sequence is human SCR-1
CC related protein.
XX
XX
SQ Sequence 292 AA;
XX
Query Match 97.1%; Score 1472; DB 22; Length 292;
Best Local Similarity 100.0%; Pred. No. 1e-108;
Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY 3 HLRILSWLFIILNMEYIGSONASRGRRORRHPNVSGCGCATGSDYNGLSCKRPL 62
DB 2 HLRILSWLFIILNMEYIGSONASRGRRORRHPNVSGCGCATGSDYNGLSCKRPL 61
QY 63 FFALRIGMKOIGVCLSSCPGSGYGRYPDINKCTKACADCTCFNNKFCFKKSGFYH 122
DB 62 FFALRIGMKOIGVCLSSCPGSGYGRYPDINKCTKACADCTCFNNKFCFKKSGFYH 121

QY 123 LGKCLDNCPEGLNANNHMECVSIHCEVSENNPWSPTCKKGTGFRGTETRVREIIQ 182
 DB 122 LGKCLDNCPEGLNANNHMECVSIHCEVSENNPWSPTCKKGTGFRGTETRVREIIQ 181
 QY 183 HPSAKGNLCPTNETRKTCTVORRKCCKGGRGKGRERKPKNGESKEAIPDSKSLSS 242
 DB 182 HPSAKGNLCPTNETRKTCTVORRKCCKGGRGKGRERKPKNGESKEAIPDSKSLSS 241
 QY 243 KEIPEORENKQOQKKRKYODKOKS 266
 DB 242 KEIPEORENKQOQKKRKYODKOKS 265

RESULT 9
 ID ABP61846 standard; Protein; 292 AA.
 AC ABP61846;
 XX
 XX 04-OCT-2002 (first entry)
 DE Human polypeptide SEQ ID NO 200.
 XX
 XX Human; cytosol; antirheumatic; antiarthritic; vulnery; analgesic;
 KW antiinflammatory; antibacterial; immunosuppressive; antiparkinsonian;
 KW neuroprotective; nocotropic; osteopathic; haemostatic; vasotropic;
 KW antitumor; fungicide; antidiabetic; antiallergic;
 KW immunostimulant; antiparasitic; secreted protein; transmembrane protein;
 KW cytokine; cell proliferation; cell differentiation; autoimmune disease;
 KW stem cell; growth factor; nervous system disease; neuropathy;
 KW Alzheimer's disease; Parkinson's disease; Huntington's disease;
 KW osteoporosis; severe combined immunodeficiency; SCID; infection;
 KW multiple sclerosis; rheumatoid arthritis; gene therapy.
 XX
 XX Homo sapiens.
 OS
 XX
 XX US200206394-A1.
 PN
 XX 30-MAY-2002.
 PD
 XX 22-DEC-2000; 2000US-0745763.
 PF
 XX 18-MAR-1998; 98US-0040963.
 PR
 XX (JACO/) JACOBS K.
 PA (MCCO/) MCCOY J.M.
 PA (LAVA/) LAVALLIE E.R.
 PA (COLL/) COLLINS-RACIE L.A.
 PA (EVAN/) EVANS C.
 PA (MERB/) MERBERG D.
 PA (TREA/) TREACY M.
 PA (SPAU/) SPAULDING V.
 XX
 XX
 PI Jacobs K, MCCOY JM, Lavallie ER, Collins-Racie LA, Evans C;
 PI Metzberg D, Treacy M, Spaulding V;
 DR WPI; 2002-582343/62.
 DR N-PSDB; ABO92060.
 DR
 XX Novel secreted or transmembrane protein and polynucleotide encoding the
 PT protein, useful for diagnosis and treatment of neurological disorders,
 PT cancer, autoimmune diseases, bone disorders and lung or liver fibrosis
 PT
 XX
 XX Claim 207, Page 203-204; 284pp; English.
 PS
 XX The invention relates to human secreted or transmembrane protein (I),
 CC their fragments and is encoded by specific complementary deoxyribonucleic
 CC acid (CDNA) inserts (II), where the protein is substantially free from
 CC other mammalian proteins. (I) are useful for preventing, treating or
 CC ameliorating a medical condition, especially immunological treatment or
 CC prevention of tumours. (I) exhibits activity relating to angiogenesis.

CC cytokine, cell proliferation, cell differentiation, antiinflammatory,
 CC stem cell growth factor activity and activin or inhibin-related
 CC activities. (I) can be used to manipulate stem cells in culture to give
 CC rise to neuroepithelial cells that can be used to augment or replace
 CC cells damaged by illness, autoimmune disease, accidental damage or
 CC genetic disorders. (I) induces the proliferation of neural cells and
 CC regeneration of nerve and brain tissue and is useful for the treatment of
 CC central and peripheral nervous system diseases and neuropathies, such as
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic
 CC lateral sclerosis. (II) is involved in chemotactic or chemokinetic
 CC activity, regulation of hematopoiesis and is useful for treating myeloid
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
 CC tissue growth and in tissue repair, healing of burns, incisions, ulcers,
 CC for treating osteoporosis, osteoarthritis, bone degenerative disorders or
 CC periodontal disease. (I) is also useful for gut protection or
 CC regeneration and treatment of lung or liver fibrosis, reperfusion injury
 CC in various tissues, various immune deficiencies and disorders including
 CC severe combined immunodeficiency (SCID), bacterial or fungal infections,
 CC autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis,
 CC diabetes mellitus, myasthenia gravis, allergic reactions and conditions,
 CC such as asthma or other respiratory problems. (II) is useful to express
 CC recombinant protein, as markers for tissues in which the corresponding
 CC protein is preferentially expressed and in gene therapy. The present
 CC sequence is that of a polypeptide of the invention.
 CC
 XX
 50 Sequence 292 AA.
 Query Match 97.1%; Score 1472; DB 23; Length 292;
 Best Local Similarity 100.0%; Pred. No. 1e-108;
 Matches 264; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 3 HLLISLWLIILNFMVEYIGSONASRGRRORRMHPNWSQCGCGCATCSDYNGCLSCPKRL 62
 DB 2 HLLISLWLIILNFMVEYIGSONASRGRRORRMHPNWSQCGCGCATCSDYNGCLSCPKRL 61
 QY 63 FFLERIRGKQIGVCLSCPSGYGTRYPDINKCTKADDDTCNNPFCCKSGGYLH 122
 DB 62 FFLERIRGKQIGVCLSCPSGYGTRYPDINKCTKADDDTCNNPFCCKSGGYLH 121
 QY 123 LGKCLDNCPEGLNANNHMECVSIHCEVSENNPWSPTCKKGTGFRGTETRVREIIQ 182
 DB 122 LGKCLDNCPEGLNANNHMECVSIHCEVSENNPWSPTCKKGTGFRGTETRVREIIQ 181
 QY 183 HPSAKGNLCPTNETRKTCTVORRKCCKGGRGKGRERKPKNGESKEAIPDSKSLSS 242
 DB 182 HPSAKGNLCPTNETRKTCTVORRKCCKGGRGKGRERKPKNGESKEAIPDSKSLSS 241
 QY 243 KEIPEORENKQOQKKRKYODKOKS 266
 DB 242 KEIPEORENKQOQKKRKYODKOKS 265

RESULT 10
 ID AAB93875 standard; Protein; 292 AA.
 AC AAB93875;
 XX
 XX 26-JUN-2001 (first entry)
 DE Human protein sequence SEQ ID NO:13761.
 XX
 XX Human; primer; detection; diagnosis; antisense therapy; gene therapy.
 XX
 XX Homo sapiens.
 OS
 XX
 XX EP1074617-A2.
 PN
 XX 07-FEB-2001.
 PD
 XX 28-JUL-2000; 2000EP-0116126.
 PF
 XX

PR 29-JUL-1999; 99JP-0248036.
 PR 27-AUG-1999; 99JP-0300253.
 PR 11-JAN-2000; 2000JP-0118776.
 PR 02-MAY-2000; 2000JP-0183767.
 PR 09-JUN-2000; 2000JP-0241899.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Isegai T, Nishikawa T, Hayashi K, Saito K, Yamamoto J;
 PI Ishii S, Sugiyama T, Wakamatsu A, Nagai K, Otsuki T;
 XX
 DR WPI; 2001-318749/34.
 XX
 PT Primer sets for synthesizing polynucleotides, particularly the 5602
 PT full-length cDNAs defined in the specification, and for the detection
 PT and/or diagnosis of the abnormality of the proteins encoded by the
 PT full-length cDNAs.
 XX
 PS Claim 8; SEQ ID 13781; 2537pp + CD ROM; English.
 XX
 CC The present invention describes primer sets for synthesizing 5602
 CC full-length cDNAs defined in the specification. Where a primer set
 CC comprises: (a) an oligo-dT primer and an oligonucleotide complementary
 CC to the complementary strand of a polynucleotide which comprises one of
 CC the 5602 nucleotide sequences defined in the specification, where the
 CC oligonucleotide comprises at least 15 nucleotides; or (b) a combination
 CC of an oligonucleotide comprising a sequence complementary to the
 CC complementary strand of a polynucleotide which comprises a 5'-end
 CC sequence and an oligonucleotide comprising a sequence complementary to a
 CC polynucleotide which comprises at least 15 nucleotides and the combination
 CC of the 5'-end sequence/3'-end sequence is selected from those defined in
 CC the specification. The primer sets can be used in antisense therapy and
 CC in gene therapy. The primers are useful for synthesizing polynucleotides,
 CC particularly full-length cDNAs. The primers are also useful for the
 CC detection and/or diagnosis of the abnormality of the proteins encoded by
 CC the full-length cDNAs. The primers allow obtaining of the full-length
 CC cDNAs easily without any specialised methods. AAH03166 to AAH13628 and
 CC AAH13633 to AAH18742 represent human cDNA sequences; AAB92446 to
 CC AAB95893 represent human amino acid sequences; and AAH13629 to AAH13632
 CC represent oligonucleotides, all of which are used in the exemplification
 CC of the present invention.
 XX
 SO Sequence 292 AA;
 Query Match 95.8%; Score 1452; DB 22; Length 292;
 Best Local Similarity 99.2%; Pred. No. 3.8e-107;
 Matches 262; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 3 HRLISMLFIILNMEYIGSONSRGRORRMRHNPVSGCGGATGSDYNGLSCKPRL 62
 DB 2 HRLISMLFIILNMEYIGSONSRGRORRMRHNPVSGCGGATGSDYNGLSCKPRL 61
 QY 63 PALERIGMKOIGVCLSSCPGYYGTRYPDINKCTKCAADCTCFPNKPFCTKCKSGFYIH 122
 DB 62 PALERIGMKOIGVCLSSCPGYYGTRYPDINKCTKCAADCTCFPNKPFCTKCKSGFYIH 121
 QY 123 LCKCLDNCPEGLAEANHTMCEVSIHVCESEWNPWSPCTKKGKTCGFKGTETREVAEIIQ 182
 DB 122 LCKCLDNCPEGLAEANHTMCEVSIHVCESEWNPWSPCTKKGKTCGFKGTETREVAEIIQ 181
 QY 183 HSAAGNLCPTNETRCKTVQRKKCKGGRGKGRKRRKPPNKGSEKAIIPDSKLESS 242
 DB 182 HSAAGNLCPTNETRCKTVQRKKCKGGRGKGRKRRKPPNKGSEKAIIPDSKLESS 241
 QY 243 KEIPEORENKOOKKRRVODKOKS 266
 DB 242 KEIPEORENKOOKKRRVODKOKS 265

RESULT 11
 AAE13153
 IF AAE13153 standard; Protein; 251 AA.

XX
 AC AAE13153;
 XX
 DT 28-JAN-2002 (first entry)
 XX
 DE Human mature stem cell growth factor-like protein.
 XX
 KW Human; stem cell growth factor-like protein; antiinflammatory; neurotropic;
 KW neuroprotective; vulnerary; cytosolic; anticonvulsant; immunostimulant;
 KW vasotrophic; virutide; dermatological; tranquiliser; cerebroprotective;
 KW osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KW duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS;
 KW acquired immune deficiency syndrome; agammaglobulinaemia; thalassaemia;
 KW Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KW adental white matter degeneration; anaemia; neurodegenerative disease;
 KW Parkinson's disease; Alzheimer's disease; thrombocytopaenia; SCID;
 KW severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KW multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KW autoimmune pulmonary inflammation; cancer; tumour; gene therapy;
 KW SCR-1; supporting factor for the proliferation of stem cell.
 XX
 OS Homo sapiens.
 XX
 PN WO200177169-A2.
 XX
 PD 18-OCT-2001.
 XX
 PF 05-APR-2001; 2001WO-US11208.
 XX
 PR 05-APR-2000; 2000US-0543774.
 PR 28-JUN-2000; 2000US-215733P.
 PR 09-JAN-2001; 2001US-0752562.
 PR 05-FEB-2001; 2001US-266614P.
 XX
 PA (HSE-) HSEBQ INC.
 PA (KIRI) KIRIN BEER KK.
 PI Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX
 DR WPI; 2001-657166/75.
 XX
 PT Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis
 XX
 PS Claim 28; Page 216-217; 232pp; English.
 CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell
 CC growth factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety
 CC of diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinaemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassaemia, haemolytic anaemia due to enzyme defect, congenital
 CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
 CC storage diseases such as mucopolysaccharidosis, adental white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the
 CC invention are useful for treating diseases such as Parkinson's
 CC disease, Alzheimer's disease and other neurodegenerative diseases,
 CC thrombocytopaenia, immune deficiencies and disorders such as severe
 CC combined immunodeficiency (SCID) and autoimmune disorders such as
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC and autoimmune pulmonary inflammation. Sequences of the invention are
 CC also useful in gene therapy. The present sequence is the mature protein
 CC of human stem cell growth factor-like protein.

XX Sequence 251 AA; 92.3%; Score 1400; DB 22; Length 251;
 SO Query Match Best Local Similarity 100.0%; Pred. No. 4.2e-103; Matches 251; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 QNASRRGRORRMHPNVSGCGGCGATCSIDYNGCLSCRPFLFALERTIGMKOIGVCLSSCP 82
 DB 1 QNASRRGRORRMHPNVSGCGGCGATCSIDYNGCLSCRPFLFALERTIGMKOIGVCLSSCP 60
 QY 83 SGYGRYPDINKCTCKACDDCTCFNNKCTCKSGFYHLGKCLDNCPESLEANNHTME 142
 DB 61 SGYGRYPDINKCTCKACDDCTCFNNKCTCKSGFYHLGKCLDNCPESLEANNHTME 120
 QY 143 CVSIHVCEVSEWNPSPCTKXGKTGCFKGTETRVAEIIOHPSAKNLCPTNETRKTCTV 202
 DB 121 CVSIHVCEVSEWNPSPCTKXGKTGCFKGTETRVAEIIOHPSAKNLCPTNETRKTCTV 180
 QY 203 ORKKCKGRGKGRERKXKPNKGSKEAIPDSKLSSESKETIPORBNKQOKKRVOD 262
 DB 181 ORKKCKGRGKGRERKXKPNKGSKEAIPDSKLSSESKETIPORBNKQOKKRVOD 240
 QY 263 KOKSVSVSTVH 273
 DB 241 KOKSVSVSTVH 251

RESULT 12
 AAE13167
 ID AAE13167 standard; Protein; 279 AA.
 XX AAE13167;
 AC AAE13167;
 XX 28-JAN-2002 (first entry)
 DT 28-JAN-2002 (first entry)
 DE Mouse stem cell growth factor-like protein.
 XX Mouse stem cell growth factor-like protein.
 XX Mouse; stem cell growth factor-like protein; antiinflammatory; neutrotropic; neuroprotective; vulnerary; cytoskeletal; anticonvulsant; immunostimulant; vasotropic; virucide; dermatological; tranquilliser; cerebroprotective; osteopathic; immunodeficiency syndrome; chronic granulomatous disease; duplicated immunodeficiency syndrome; Wiskott-Aldrich syndrome; AIDS; acquired immune deficiency syndrome; agammaglobulinemia; thalassemia; Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis; adrenal white matter degeneration; anaemia; neurodegenerative disease; Parkinson's disease; Alzheimer's disease; thrombocytopenia; SCID; severe combined immunodeficiency; immune disorder; autoimmune disorder; multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis; autoimmune pulmonary inflammation; cancer; tumour; gene therapy; SCR-1; supporting factor for the proliferation of stem cell.
 XX Mus musculus.
 OS Mus musculus.
 XX Key Location/Qualifiers
 FH Peptide 1..21
 FT /label= Signal_peptide
 FT Protein 22..279
 FT /note= "Mouse mature stem cell growth factor-like protein"
 FT factor-like protein"
 XX WO200177169-A2.
 XX 18-OCT-2001.
 PD 18-OCT-2001.
 XX 05-APR-2001; 2001MO-US11208.
 PF 05-APR-2001; 2001MO-US11208.
 XX 05-APR-2000; 2000US-0543774.
 PR 28-JUN-2000; 2000US-215733P.
 PR 09-JAN-2001; 2001US-0757562.
 PR 05-FEB-2001; 2001US-266614P.
 XX (HYSE-) HYSEQ INC.

PA (KIRI) KIRIN BEER KK.
 XX Tang TY, Iabat I, Tillinghaas JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Grain B, Dickson W, Mize NK, Nishikawa M;
 XX WPI; 2001-657166/75.
 DR N-PSDB; NAD21727.
 XX Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis
 PS Claim 28; Page 223-224; 232pp; English.

CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferentially
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell
 CC growth factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety
 CC of diseases such as immunodeficiency syndrome, chronic granulomatous
 CC disease, duplicated immunodeficiency syndrome, agammaglobulinemia,
 CC Wiskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
 CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
 CC storage diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the
 CC invention are useful for treating diseases such as Parkinson's
 CC disease, Alzheimer's disease and other neurodegenerative diseases,
 CC thrombocytopenia, immune deficiencies and disorders such as severe
 CC combined immunodeficiency (SCID) and autoimmune disorders such as
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC and autoimmune pulmonary inflammation. Sequences of the invention are
 CC also useful in gene therapy. The present sequence is stem cell growth
 CC factor-like protein from mouse.

XX Sequence 279 AA;
 SO Query Match 86.4%; Score 1310.5; DB 22; Length 279;
 Best Local Similarity 87.1%; Pred. No. 5.8e-96;
 Matches 242; Conservative 11; Mismatches 18; Indels 7; Gaps 4;

QY 3 HLRLISLFTILNFMFYISQNASRRGRORRMHPNVSGCGGCGATCSIDYNGCLSCRPRL 62
 DB 2 HLRLISLFTILNFMFYISQNASRRGRORRMHPNVSGCGGCGATCSIDYNGCLSCRPRL 61
 QY 63 FFALERIGMKOIGVCLSSCPSGYVSTRYPDINKCTCKACDDCTCFNNKCTCKSGFYHL 122
 DB 62 FFALERIGMKOIGVCLSSCPSGYVSTRYPDINKCTCKACDDCTCFNNKCTCKSGFYHL 121
 QY 123 LGKCLDNCPESLEANNHTMECVSIHVCEVSEWNPSPCTKXGKTGCFKGTETRVAEIIO 182
 DB 122 LGKCLDNCPESLEANNHTMECVSIHVCEVSEWNPSPCTKXGKTGCFKGTETRVAEIIO 181
 QY 183 HPSA-KGNLCPTNETRKTCTVORKKCKGRGKGRERKXKPNKGSKE--AIPDSKS 238
 DB 182 HPSAKGKGNLCPTNETRKTCTVORKKCKGRGKGRERKXKPNKGSKE--AIPDSKS 241
 QY 239 LSSKEIPEORENK--QOQKRRVOD--OKSVSVSTVH 273
 DB 242 LSSKEIPEORENK--QOQKRRVOD--OKSVSVSTVH 279

RESULT 13
 AAE13149
 ID AAE13149 standard; Protein; 160 AA.
 XX AAE13149;
 AC AAE13149;

DT 28-JAN-2002 (first entry)
 XX Human stem cell growth factor-like protein #1.
 XX
 KM Human; stem cell growth factor-like protein; antiinflammatory; neutrophic;
 KM neuroprotective; vulnary; cytoskeletal; anticonvulsant; immunostimulant;
 KM vasotropic; virucidal; dermatological; tranquilizer; cerebroprotective;
 KM osteopathic; immunodeficiency syndrome; chronic granulomatous disease;
 KM duplicated immunodeficiency syndrome; Miskott-Aldrich syndrome; AIDS;
 KM acquired immune deficiency syndrome; agammaglobulinemia; thalassemia;
 KM Gaucher's disease; lysosomal storage disease; mucopolysaccharidosis;
 KM adrenal white matter degeneration; anaemia; neurodegenerative disease;
 KM Parkinson's disease; Alzheimer's disease; thrombocytopenia; SCID;
 KM severe combined immunodeficiency; immune disorder; autoimmune disorder;
 KM multiple sclerosis; systemic lupus erythematosus; rheumatoid arthritis;
 KM autoimmune pulmonary inflammation; cancer; tumour; gene therapy;
 KM SCR-1; supporting factor for the proliferation of stem cell.
 XX
 OS Homo sapiens.
 XX
 PN WO200177169-A2.
 XX
 XX 18-OCT-2001.
 PD
 XX 05-APR-2001; 2001WO-US11208.
 PF
 XX 05-APR-2000; 2000US-0543774.
 PR 28-JUN-2000; 2000US-215733P.
 PR 09-JAN-2001; 2001US-0757562.
 PR 05-FEB-2001; 2001US-266614P.
 XX
 PA (HYSE-) HYSEO INC.
 PA (KIRI) KIRIN BEER KK.
 XX
 PI Tang TY, Labat I, Tillinghast JS, Sinku A, Liu C, Drmanac RT;
 PI Stache-Crain B, Dickson M, Mize NK, Nishikawa M;
 XX
 DR WPI: 2001-657166/75.
 DR N-PSDB: AAD21723.
 XX
 PT Novel stem cell growth factor like polypeptides and polynucleotides for
 PT identifying modulators useful for treating diseases such as Alzheimer's
 PT disease, cancer, rheumatoid arthritis, osteoporosis
 XX
 PS Example 2; Page 166; 232pp; English.
 XX
 CC The patent discloses novel stem cell growth factor-like proteins and
 CC polynucleotides encoding them. Proteins of the invention are also known
 CC as supporting factor for the proliferation of stem cells (SCR-1). Stem
 CC cell growth factor-like proteins are useful for supporting proliferation
 CC or survival of a stem cell or germ cell which is preferably primordial
 CC germ cell, germ line stem cell, embryonic stem cell, haematopoietic stem
 CC cell, haematopoietic progenitor cell, pluripotent cell or totipotent
 CC cell. The haematopoietic progenitor cell cultured using stem cell
 CC growth factor-like proteins can replace as a graft for the bone marrow
 CC transplantation or cord blood transplantation for treating a variety
 CC of diseases such as immunodeficiency syndrome, agammaglobulinemia,
 CC disease, duplicated immunodeficiency syndrome, chronic granulomatous
 CC disease, Miskott-Aldrich syndrome, acquired immune deficiency syndrome (AIDS),
 CC thalassemia, haemolytic anaemia due to enzyme defect, congenital
 CC anaemia such as sickle cell anaemia, Gaucher's disease, lysosomal
 CC storage diseases such as mucopolysaccharidosis, adrenal white matter
 CC degeneration, a variety of cancer and tumours. Proteins of the
 CC invention are useful for treating diseases such as Parkinson's
 CC disease, Alzheimer's disease and other neurodegenerative diseases,
 CC thrombocytopenia, immune deficiencies and disorders such as severe
 CC combined immunodeficiency (SCID) and autoimmune disorders such as
 CC multiple sclerosis, systemic lupus erythematosus, rheumatoid arthritis,
 CC and autoimmune pulmonary inflammation. Sequences of the invention are
 CC also useful in gene therapy. The present sequence is stem cell growth
 CC factor-like protein from human.
 XX
 S2 Sequence 160 AA:

Query Match 59.6%; Score 903; DB 22; Length 160;
 Best Local Similarity 100.0%; Pred. No. 5,8e-64;
 Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 96 CTCKACDPCFPKNCCTCKSGFYLHIGKLDNCPGELFANNHTECVSIYHCEVSEWN 155
 DB 1 CTCKACDPCFPKNCCTCKSGFYLHIGKLDNCPGELFANNHTECVSIYHCEVSEWN 60
 QY 156 FMSPTCKKGTCTGFKGTETRVREIIHQPSAKGNLCPTNETRKTCTVQRKCKQKERGSK 215
 DB 61 FMSPTCKKGTCTGFKGTETRVREIIHQPSAKGNLCPTNETRKTCTVQRKCKQKERGSK 120
 QY 216 GRERRKKPKNGESKEAIPDSKLSSESKETPEORENKKQ 255
 DB 121 GRERRKKPKNGESKEAIPDSKLSSESKETPEORENKKQ 160
 RESULT 14
 ABB1374
 ID ABB1374 standard; peptide; 160 AA.
 XX
 AC ABB1374;
 DT 11-JAN-2002 (first entry)
 XX
 DE Human secreted protein homologue, SEQ ID NO:1744.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; growth factor;
 KM haematopoiesis regulation; tissue growth; immunomodulator; activin;
 KM inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
 KM proliferation; metastasis; cancer; tumour; haematopoietic disorder;
 KM myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
 KM chronic inflammatory condition; proliferative retinopathy;
 KM atherosclerosis; coronary heart disease; arterial ischaemia;
 KM bone disorder; osteoporosis; vascular growth disorder;
 KM tissue regeneration; wound healing; gene therapy; immune disorder;
 KM cell culture; drug screening; infection; antiinflammatory;
 KM antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
 KM cytoskeletal; osteopathic; vasotropic; cardiant; virucide; antibacterial;
 KM antifungal; vulnary; antulcer.
 XX
 OS Homo sapiens.
 XX
 XX WO200157188-A2.
 XX
 PD 09-AUG-2001.
 XX
 PF 05-FEB-2001; 2001WO-US03800.
 XX
 PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 XX
 PA (HYSE-) HYSEO INC.
 XX
 PI Tang YT, Liu C, Drmanac RT;
 PI WPI: 2001-457740/49.
 DR N-PSDB: ABA08618.
 XX
 PT Human proteins and DNA encoding sequences useful for preventing
 PT treating or ameliorating a medical condition in a mammalian subject
 PT e.g. arthritis and cancer.
 XX
 PS Claim 20; Page 174-175; 1963pp; English.
 XX
 CC Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
 CC sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
 CC invention also relates to vectors and recombinant host cells comprising a
 CC nucleotide of the invention, methods of producing the novel polypeptides,
 CC antibodies against the polypeptides, methods of detecting the nucleotides
 CC or polypeptides in a sample, and methods of identifying compounds which
 CC bind to polypeptides of the invention. Although novel, many of the

CC polypeptides of the invention have homology to known proteins, thereby
 CC giving an insight into their probable biological activities, and hence
 CC potential therapeutic applications. The polypeptides of the invention may
 CC have various activities, including cytokine, cell proliferation or cell
 CC differentiation activities; stem cell growth factor activity;
 CC haematopoietic regulatory activity; tissue growth activity;
 CC immunomodulatory activity; activin- or inhibin-related activities;
 CC chemotactic or chemokinetic activities; haemostatic, thrombotic or
 CC thrombolytic activities; receptor or ligand activities; or may be
 CC involved in oncogenesis, cancer cell proliferation or metastasis.
 CC Depending on their biological activities, polypeptides and nucleotides of
 CC the invention are useful for preventing, treating or ameliorating medical
 CC conditions, e.g., by protein or gene therapy. Such conditions include
 CC cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
 CC disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
 CC proliferative retinopathy, atherosclerosis, coronary heart disease,
 CC arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
 CC vascular growth. Polypeptides involved with tissue regeneration and
 CC repair (or nucleic acids encoding them) may be used to promote wound
 CC healing (e.g., of burns, incisions and ulcers), while those with
 CC immunomodulatory activities may be used in the treatment of viral,
 CC bacterial and fungal infections in addition to immune disorders.
 CC Polypeptides with growth factor activity may be used in cell cultures to
 CC promote cell growth. For example, such polypeptides may be used to
 CC manipulate stem cells in culture to give rise to neuroepithelial cells
 CC that can be used to augment or replace cells damaged by illness.
 CC autoimmune disease or accidental damage. The polypeptides and nucleotides
 CC may also be used in the diagnosis of the above conditions, and in drug
 CC screening techniques. The present sequence represents a novel human
 CC polypeptide of the invention.
 CC
 CC Sequence 160 AA;
 SQ
 Query Match 59.6%; Score 903; DB 22; Length 160;
 Best Local Similarity 100.0%; Pred. No. 5.8e-64;
 Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 96 CTCKADCDTCTFNKPKCTCKSGFYHLGKCLDNCPEGLFANNHTMECVSIHCEVSEMN 155
 DB 1 CTCKADCDTCTFNKPKCTCKSGFYHLGKCLDNCPEGLFANNHTMECVSIHCEVSEMN 60
 QY 156 PWSPTCKGKTGCFRGTETRVREIIQHPSAKGNLCPTNETRKTCTVQRKCKGGERGKX 215
 DB 61 PWSPTCKGKTGCFRGTETRVREIIQHPSAKGNLCPTNETRKTCTVQRKCKGGERGKX 120
 QY 216 GREKRRKPKNGESKEAIPDSKSLSSKEIPEORENKQOO 255
 DB 121 GREKRRKPKNGESKEAIPDSKSLSSKEIPEORENKQOO 160
 RESULT 15
 AAM79312
 ID AAM79312 standard; Protein; 160 AA.
 AC AAM79312;
 DT 06-NOV-2001 (first entry)
 DE Human protein SEQ ID NO 2958.
 XX
 XX Human; cytokine; cell proliferation; cell differentiation; gene therapy;
 KW tissue growth factor; immunomodulatory; cancer; leukaemia;
 KM nervous system disorder; arthritis; inflammation.
 XX
 XX Homo sapiens.
 XX
 XX WO200157190-A2.
 XX
 XX 09-AUG-2001.
 XX
 XX 05-FEB-2001; 2001MO-US04098.
 XX

PR 03-FEB-2000; 2000US-0496914.
 PR 27-APR-2000; 2000US-0560875.
 PR 20-JUN-2000; 2000US-0598075.
 PR 19-JUL-2000; 2000US-0620325.
 PR 01-SEP-2000; 2000US-0654936.
 PR 15-SEP-2000; 2000US-0663561.
 PR 20-OCT-2000; 2000US-0693125.
 PR 30-NOV-2000; 2000US-0728422.
 XX
 XX (HSE-) HYSEQ INC.
 XX
 XX Tang YT, Liu C, Dumanac RT, Asundi V, Zhou P, Xu C, Cao Y, Ma Y,
 PI Zhao QA, Wang D, Wang J, Zhang J, Ren F, Chen R, Wang ZW;
 PI Xue AJ, Yang Y, Wejhtman T, Goodrich R;
 DR WPI: 2001-476283/51.
 DR N-PSDB; AAK52445.
 XX
 PT Nucleic acids encoding polypeptides with cytokine-like activities,
 PT useful in diagnosis and gene therapy -
 PS
 PS Claim 20; Page 214; 6221pp; English.
 XX
 XX The invention relates to polynucleotides (AAK51456-AAK53435) and the
 CC encoded polypeptides (AAM78123-AAK80302) that exhibit activity elating to
 CC cytokine, cell proliferation or cell differentiation or which may induce
 CC production of other cytokines in other cell populations. The
 CC polynucleotides and polypeptides are useful in gene therapy, vaccines or
 CC peptide therapy. The polypeptides have various cytokine-like activities,
 CC e.g. stem cell growth factor activity, haematopoietic regulatory
 CC activity, tissue growth factor activity, immunomodulatory activity and
 CC activin/inhibin activity and may be useful in the diagnosis and/or
 CC treatment of cancer, leukaemia, nervous system disorders, arthritis and
 CC inflammation.
 CC Note: Records for SEQ ID NO 2110 (AAK52581), 2111 (AAK52582) and 3666
 CC (AAM90020) are omitted as the relevant pages from the sequence listing
 CC were missing at the time of publication.
 CC
 CC Sequence 160 AA;
 SQ
 Query Match 59.6%; Score 903; DB 22; Length 160;
 Best Local Similarity 100.0%; Pred. No. 5.8e-64;
 Matches 160; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 96 CTCKADCDTCTFNKPKCTCKSGFYHLGKCLDNCPEGLFANNHTMECVSIHCEVSEMN 155
 DB 1 CTCKADCDTCTFNKPKCTCKSGFYHLGKCLDNCPEGLFANNHTMECVSIHCEVSEMN 60
 QY 156 PWSPTCKGKTGCFRGTETRVREIIQHPSAKGNLCPTNETRKTCTVQRKCKGGERGKX 215
 DB 61 PWSPTCKGKTGCFRGTETRVREIIQHPSAKGNLCPTNETRKTCTVQRKCKGGERGKX 120
 QY 216 GREKRRKPKNGESKEAIPDSKSLSSKEIPEORENKQOO 255
 DB 121 GREKRRKPKNGESKEAIPDSKSLSSKEIPEORENKQOO 160
 Search completed: May 6, 2003, 14:51:39
 Job time : 38 secs

RT morphogenetic proteins at distinct sites during embryogenesis.";
 RL J. Cell Biol. 134:181-191 (1996).
 RN (6)
 RP DEVELOPMENTAL EXPRESSION
 RX MEDLINE=97436919; PubMed=9291583;
 RA Rancourt S.L., Rancourt D.E.;
 RT "Murine subtilisin-like proteinase SPC6 is expressed during embryonic
 implantation, somitogenesis, and skeletal formation.";
 RL Dev. Genet. 21:75-81 (1997).
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPROTEASE ACTIVITY
 WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF. MAY BE RESPONSIBLE
 FOR THE MATURATION OF GASTROINTESTINAL PEPTIDES. MAY BE INVOLVED
 IN THE CELLULAR PROLIFERATION OF ADRENAL CORTEX VIA THE ACTIVATION
 OF GROWTH FACTORS.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 PROTEINASES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
 CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC SUBCELLULAR LOCATION: PCSA IS SECRETED THROUGH THE REGULATED
 SECRETORY PATHWAY. PCSB IS A TYPE I MEMBRANE PROTEIN LOCALIZED TO
 A PARANUCLEAR POST-GOLGI NETWORK COMPARTMENT IN COMMUNICATION WITH
 EARLY ENDOPLASM.
 CC -1- ALTERNATIVE PRODUCTS: AT LEAST 2 ISOFORMS: PCSB/LONG (SHOWN HERE)
 AND PCSA/SHORT: ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: PCSA IS EXPRESSED IN MOST TISSUES, BUT IS MOST
 ABUNDANT IN THE INTESTINE AND ADRENALS. PCSB IS EXPRESSED IN THE
 INTESTINE, ADRENALS AND LUNG BUT NOT IN THE BRAIN.
 CC -1- DEVELOPMENTAL STAGE: WEAKLY EXPRESSED THROUGHOUT THE EMBRYO,
 BUT MARKEDLY UPREGULATED AT DISCRETE SITES DURING DEVELOPMENT. AT
 E6.5, PROMINENT EXPRESSION OBSERVED IN DIFFERENTIATED DECIDUA. AT
 E7.5, INTENSE EXPRESSION IN EXTRAMERITONIC ENDODERM, AMNION AND
 NASCENT MESODERM. AT 8.5, ABUNDANT EXPRESSION IN SOMITES AND YOLK
 SAC FOLLOWED BY A CONFINATION TO DERMATOTOME COMPARTMENT. BETWEEN
 E9.5 AND E11.5, ABUNDANT EXPRESSION IN AER (THICKENED ECTODERMAL
 CELLS OF LIMB BUDS). AT E12.5, EXPRESSION IN THE LIMBS IS CONFINED
 TO THE CONDENSING MESENCHYM SURROUNDING THE CARILAGE. AT THIS
 STAGE, STRONG EXPRESSION ALSO DETECTED IN VERTEBRAL AND FACIAL
 CARILAGE PRIMORDIA AND IN THE MUSCLE OF THE TONGUE. AT E16.5,
 ABUNDANT EXPRESSION IN EPITHELIAL CELLS OF THE INTESTINAL VILLI.
 CC ISOFORM A IS MOST ABUNDANT AT ALL STAGES BUT SIGNIFICANT LEVELS OF
 ISOFORM B OCCUR AT E12.5.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
 RETICULUM.
 CC -1- DOMAIN: AC 1 AND AC 2 (CLUSTERS OF ACIDIC AMINO ACIDS) CONTAIN
 SORTING INFORMATION. AC 1 DIRECTS TGN LOCALIZATION AND INTERACTS
 WITH THE TGN SORTING PROTEIN PACS-1.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; D17583; BA04507.1; -;
 DR EMBL; D12619; BA002143.1; -;
 DR EMBL; L14932; AA074636.1; -;
 DR PIR; JX0248; JX0248.
 DR PIR; A48225; A48225.
 DR HSSP; G99405; IMPT.
 DR MEROPS; S08.076; -;
 DR MGD; MG1.97515; PGSK5.
 DR InterPro; IPR000561; EGF-like.
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002884; P_domain.
 DR InterPro; IPR000209; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF01483; P_PARTIAL.

DR PRINTS; PR00723; SUBTILISIN.
 DR Prodom; PD000717; P_domain; 1.
 DR SMART; SM00181; EGF_3.
 DR SMART; SM00001; EGF_like; 2.
 DR SMART; SM00261; FU_-22.
 DR PROSITE; PS00136; SUBTILASE ASP; 1.
 DR PROSITE; PS00137; SUBTILASE_HIS; 1.
 DR PROSITE; PS00138; SUBTILASE_SER; 1.
 DR Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 KM Cleavage on pair of basic residues; Repeat; Alternative splicing;
 KW Transmembrane.
 FT SIGNAL 1 34
 FT PROPEP 35 116
 FT CHAIN 117 1877
 FT
 FT DOMAIN 117 1768
 FT TRANSMEM 1769 1789
 FT DOMAIN 1790 1877
 FT DOMAIN 117 452
 FT DOMAIN 464 602
 FT DOMAIN 638 1753
 FT DOMAIN 1825 1844
 FT DOMAIN 1856 1877
 FT SITE 116 117
 FT SITE 521 523
 FT ACT_SITE 173 173
 FT ACT_SITE 214 214
 FT ACT_SITE 388 388
 FT CARBOHYD 227 227
 FT CARBOHYD 383 383
 FT CARBOHYD 667 667
 FT CARBOHYD 754 754
 FT CARBOHYD 804 804
 FT CARBOHYD 854 854
 FT CARBOHYD 951 951
 FT CARBOHYD 1016 1016
 FT CARBOHYD 1220 1220
 FT CARBOHYD 1317 1317
 FT CARBOHYD 1523 1523
 FT CARBOHYD 1711 1711
 FT CARBOHYD 1733 1733
 FT VARSPPLIC 878 915
 FT
 FT VARSPPLIC 916 1877
 FT SEQUENCE 1877 AA; 209287 MW; EC85DE2DF2DEA1C3 CRC64;
 SQ
 Query Match 12.3%; Score 186.5; DB 1; Length 1877;
 Best Local Similarity 25.0%; Pred. No. 7.3e-06;
 Matches 66; Conservative 31; Mismatches 90; Indels 77; Gaps 15;
 QY 39 SGGCGGCGATCSGYNQ--CLGCKPFLPFLERIGMKOIGVCLSSGCSGYGRYPD--N 94
 DB 1209 NQPHSSCKTC--NSGLCASCPGTOMYLMD--ACVPSCPG--TWSPVTSG 1253
 QY 95 KCTCKKADCDCTCNKFKCTCKS--GFYLHLGKCLDNCPEGLANNTMR-CVSIYHC 149
 DB 1254 SCENCSDDCVSGSGADLCQCGLSQPDNTLLHBCRCYHSPBGFYAKDGVCHSS--PC 1311
 QY 150 EVSEMNPMSPCTKKG--KTC--GFKRGTRVRELIQHPBSKGL- 190
 DB 1312 KTCGNATSCNSCEGDFVLDHGVCKTCKPEKGVAVBVCVCHGCRPODILHKTCKCMP 1371
 QY 191 -----CPT--NETRKCTVQRKCKCKGKRGKGRKRRKPKGSKAIPDS 236
 DB 1372 DFLYNDMCHRSCPKSFYDMQCVPCCHKNC-----LECGPVEDDCKVCADTS 1420
 QY 237 KSLSS--KEIPE-----ORENKO 253
 DB 1421 KALNGLCLDECEBGTGKEEENDE 1444
 RESULT 2

PKCS_BRACL
ID PKCS_BRACL STANDARD: PRT: 1696 AA.
AC 09N15; 09N16; 09N14;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
GN (Proteinase PC6-1) (apc6).
OS Branchiostoma californiensis (California lancelet) (Amphioxus).
OC Eukaryota; Metazoa; Chordata; Cephalochordata; Branchiostomidae;
OC Branchiostoma.
OX NCBI_TaxID=7738;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS A, B AND C).
RX MEDLINE=20175281; PubMed=10708688;
RA Oliva A.A. Jr., Chan S.J., Steiner D.F.;
RT "Evolution of the prohormone convertases: identification of a
RT homolog of PC6 in the protochordate amphioxus";
RL Biochim. Biophys. Acta 1477:338-348(2000).
CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
CC OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
CC PROPEPTIDES BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS, WHERE XAA
CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
CC -1- SUBCELLULAR LOCATION: ISOFORM A AND ISOFORM C ARE SECRETED.
CC -1- ISOFORM B IS A TYPE I MEMBRANE PROTEIN.
CC -1- ALTERNATIVE PRODUCTS: 3 ISOFORMS, A, B (SHOWN HERE) AND C; ARE
CC PRODUCED BY ALTERNATIVE SPLICING.
CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
CC RETICULUM.
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY 58.
CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF184615; AAF26300.1; -;
DR EMBL; AF184616; AAF26301.1; -;
DR EMBL; AF184617; AAF26302.1; -;
DR HSP; O99405; 1MPT.
DR MEROFS; S08.0P8; -;
DR InterPro; IPR000561; EGF-like.
DR InterPro; IPR002174; Furin-like.
DR InterPro; IPR002884; P_domain.
DR InterPro; IPR002029; Peptidase_S8.
DR Pfam; PF00082; Peptidase_S8; 1.
DR Pfam; PF01483; P; 1.
DR PRINTS; PR00723; SUBTILISIN.
DR PRODOM; PD000717; P_domain; 1.
DR SMART; SM00181; EGF; 2.
DR SMART; SM00261; FU; 17.
DR PROSITE; PS00136; SUBTILASE ASP; FALSE_NEG.
DR PROSITE; PS00137; SUBTILASE HIS; 1.
DR PROSITE; PS00138; SUBTILASE SER; 1.
KW Hydrolyase: Serine protease; Glycoprotein; Zymogen; Signal;
KW Cleavage on pair of basic residues; Repeat; Alternative splicing;
KW Transmembrane.
FT SIGNAL 1 25 POTENTIAL.
FT PROPEP 26 110 POTENTIAL.
FT CHAIN 111 1696 PROPROTEIN CONVERTASE SUBTILISIN/KEXIN
FT DOMAIN 111 1618 EXTRACELLULAR (POTENTIAL).
FT DOMAIN 1619 1639 POTENTIAL.
FT DOMAIN 1640 1696 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 111 488 CATALYTIC.

FT DOMAIN 496 637 HOMO B.
FT DOMAIN 664 1649 CYS-RICH MOTIF (CRM) REGION.
FT SITE 110 111 CLEAVAGE (AUTO-) (BY SIMILARITY).
FT ACT_SITE 132 192 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 233 233 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT ACT_SITE 407 407 CHARGE RELAY SYSTEM (BY SIMILARITY).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 885 885 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARSPLIC 1259 1323 DDTIIRGECITSCGPGYEMRREKCKACHPTCKECSDEY
FT DDTCTACNDGFLITDASSCEAGCP -> ANONQASCFPA
FT PREVSVALAELARHRSYLTPTDPPQSPSPDITVLGDRRL
FT TTITSAGRCA (IN ISOFORM C).
FT MISSING (IN ISOFORM C).
FT CHPTCKECSDEYDDTCTACNDGFLITDASSCEAGCPGQFL
FT HHGDCDSCHRECKTC -> IARCVRDRRSWCDLVIRFNFC
FT VRRYFVRCCTGCKLYMEDRPMRGSSOPTQGRN (IN
FT ISOFORM A).
FT VARSPLIC 1344 1696 MISSING (IN ISOFORM A).
FT SEQUENCE 1696 AA, 108410 MM, 281CBE1784257CBD CRG64;
SO QUERY MATCH 12.24; Score 184.5; DB 1; Length 1696;
Best Local Similarity 22.74; Pred. No. 9.2e-06;
Matches 64; Conservative 25; Mismatches 74; Indels 119; Gaps 13;
OY 34 MHPNVSQGGCGGCGATCS--DYNGLSCRPRLF----- 64
DB 1328 LHHGDDSDSHRECKTCDDPHNDLSCGPGSYINDQCSHCPBETFEYEDDSGETVL 1387
OY 65 -----ALERTGMKQIGVCLSSCPSPGYTRYPPDI-NKCTKCYA 101
DB 1388 QCLCHVNCCKTCHGEGEDCMECANDIKYKQDRCVTGCEQGH--YPLDTNCCQCS 1443
OY 102 DCDTCF--NKNFCTCKSGFYHLGKLCNCEGL-----EANN 138
DB 1444 DCTCGPRNDQCVTCPPNYLVLGKLCDECEGYDTPMROKEGEGCHSCATNCGGN 1503
OY 139 HTMECHSYTH-----CEVSEW-----NMSPTCKKATCGFKGTERREI 181
DB 1504 Y--NCLSCPGSKLGEVGYCPMCEHNYVETKQICEBDNSCKTC--KSTADHCLSC 1558
OY 182 QHP--SAKGNLC-----PTNE-----TRKCTVYORK 205
DB 1559 EAPYGYAMKHLCTACCEGSPENRYCCICHESTRLCTIDRE 1600
RESULT 3
FUR2_DROME STANDARD: PRT: 1680 AA.
ID FUR2_DROME
AC P30432;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Furin-like protease 2 precursor (EC 3.4.21.75) (Furin 2).
GN FUR2
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Mandibulata; Pancrustacea; Hexapoda;
OC Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
OC Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=923181036; PubMed=1512259;
RA Roebroek A.J.M., Creemers J.W.M., Pauli I.G.L., Kurzik-Dunke U.,
RA Rentrop W., Galleff E.A.F., Leunissen J.A.M., van de Ven W.J.,
RT "Cloning and functional expression of Dufurin2, a subtilisin-like
RT proprotein processing enzyme of Drosophila melanogaster with multiple
RT repeats of a cysteine motif.";
RL J. Biol. Chem. 267:17208-17215(1992).
CC -1- FUNCTION: FURIN IS LIKELY TO REPRESENT THE UBIQUITOUS ENDOPEPTIDASE
CC ACTIVITY WITHIN CONSTITUTIVE SECRETORY PATHWAYS AND CAPABLE OF
CC CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: Release of mature proteins from their

Query Match	Best Local Similarity	11.6%	Score 176.5	DB 1	Length 1660
Matches 60	Conservative 24	Mismatches 77	Indels 53	Gaps 12	
3 HLELISLFIILINMETYIGSDNRSRRRRMRHPRVSGCGGCGCATCSY-NGCLSKRPR 61					
CC propeptides by cleavage of Arg-Xaa-Yaa-Arg- -Zaa bonds, where Xaa					
CC can be any amino acid and Yaa is Arg or Lys. Releases albinin,					
CC complement component C3 and von Willebrand factor from their					
CC respective precursors.					
CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8. FURIN SUBFAMILY.					
CC This SWISS-PROT entry is copyright. It is produced through a collabora-					
CC tion between the Swiss Institute of Bioinformatics and the EMBL outstation -					
CC the European Bioinformatics Institute. There are no restrictions on its					
CC use by non-profit institutions as long as its content is in no way					
CC modified and this statement is not removed. Usage by and for commercial					
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/					
CC or send an email to license@isb-sib.ch).					
CC					
CC EMBL, M94375, AAA28551.1; --					
CC PIR, A4343; A43434.					
CC HSP, Q99405, IMPT.					
CC MEROPS, S08.049; --					
CC FlyBase, Fggn0004598; Fur2.					
CC InterPro, IPR000561; EGF-like.					
CC InterPro, IPR002174; Furin-like.					
CC InterPro, IPR002884; P domain.					
CC InterPro, IPR000209; Peptidase_S8.					
CC Pfam, PF00082; Peptidase_S8; 1.					
CC Pfam, PF01483; P. PARTIAL					
CC PRINTS, PR00723, SUBTILISIN.					
CC Prodom, PD000717; P domain; 1.					
CC SMART, SM00181; EGF; 1.					
CC SMART, SM00261; FU; 10.					
CC PROSITE, PS00136; SUBTILASE ASP; 1.					
CC PROSITE, PS00137; SUBTILASE HIS; 1.					
CC PROSITE, PS00138; SUBTILASE SER; 1.					
CC Hydrolase; Serine protease; Glycoprotein; Signal; Transmembrane;					
CC Multigene family; Zymogen; Repeat.					
CC FT SIGNAL 1 2					
CC FT PROPEP 2 319					
CC FT CHAIN 320 1660					
CC FT ACT_SITE 418 418					
CC FT ACT_SITE 457 457					
CC FT ACT_SITE 638 638					
CC FT DOMAIN 962 1444					
CC FT REPEAT 962 1007					
CC FT REPEAT 1008 1057					
CC FT REPEAT 1058 1104					
CC FT REPEAT 1105 1153					
CC FT REPEAT 1154 1205					
CC FT REPEAT 1206 1254					
CC FT REPEAT 1255 1299					
CC FT REPEAT 1300 1346					
CC FT REPEAT 1347 1393					
CC FT REPEAT 1394 1444					
CC FT TRANSMEM 1508 1532					
CC FT DOMAIN 1533 1660					
CC FT CARBOHYD 3 3					
CC FT CARBOHYD 109 109					
CC FT CARBOHYD 130 130					
CC FT CARBOHYD 203 203					
CC FT CARBOHYD 443 443					
CC FT CARBOHYD 481 481					
CC FT CARBOHYD 928 928					
CC FT CARBOHYD 1061 1061					
CC FT CARBOHYD 1182 1182					
CC FT CARBOHYD 1275 1275					
CC FT CARBOHYD 1278 1278					
CC FT CARBOHYD 1440 1440					
CC SEQUENCE 1660 AA; 18359 MW; 0A99CEB770ABE293 CRC64;					

RESULT 4

PACC4_HUMAN	STANDARD;	PRT; 969 AA.
ID_PACC4_HUMAN		
AC P29127; Q15099; Q15100; Q9UEB1; Q9UEB2; Q9UEB7; Q9UEJ8; Q9UEJ9; Q9UEG7; Q9Y439; Q9Y4H0; Q9Y4H1; Q9Y4H2; Q9Y4G9; Q9Y4H0; Q9Y4H1;		
DC 01-DEC-1992 (Rel. 24, Created)		
DT 01-DEC-1992 (Rel. 24, Last sequence update)		
DE 15-JUN-2002 (Rel. 41, Last annotation update)		
DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)		
DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like proprotein		
DE convertase 4) (SFCA).		
DE PACE4.		
CN PACE4.		
OS Homo sapiens (Human).		
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
OX NCBI_Taxid:9606;		
RX SEQUENCE FROM N.A. (ISOFORMS PACE4A-I AND PACE4B).		
RC TISSUE=Hepatoma, and Kidney;		
RL MEDLINE=92075167; PubMed=1741956;		
RA Kiefer M.C., Tucker J.E., Joh R., Landsberg K.E., Saltman D.,		
RA Barr P.J.;		
RT "Identification of a second human subtilisin-like protease gene in		
RL the les/lfs region of chromosome 15.";		
RL DNA Cell Biol. 10:757-769(1991).		
[2]		
RX SEQUENCE FROM N.A. (ISOFORMS PACE4C AND PACE4D).		
RC TISSUE=Placenta;		
RL MEDLINE=94235049; PubMed=8179631;		
RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,		
RA Matsuda Y.;		
RT "Identification of novel cDNAs encoding human kexin-like protease,		
PACE4 isoforms.";		
RL Biochem. Biophys. Res. Commun. 200:943-950(1994).		
[3]		
RX ERRATUM.		
RL MEDLINE=95071480; PubMed=7980617;		
RA Tsuji A., Higashine K., Hine C., Mori K., Tamai Y., Nagamune H.,		
RA Matsuda Y.;		
RT "Identification of novel cDNAs encoding human kexin-like protease,		
PACE4 isoforms.";		
RL Biochem. Biophys. Res. Commun. 204:1381-1382(1994).		
[4]		
RX SEQUENCE FROM N.A. (ISOFORM PACE4A-II).		
RC TISSUE=Placenta;		
RL MORI K., Iimaki A., Kii S., Nagamune H., Nagahama M., Tsuji A.,		
RA Matsuda Y.;		
RT "Identification of a novel PACE4 isoform, PACE4A.";		
RL Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.		
[5]		
RX SEQUENCE FROM N.A. (ISOFORMS PACE4E-I AND PACE4E-II).		
RC TISSUE=Cerebellum;		
RL MEDLINE=97335942; PubMed=9192737;		
RA Mori K., Kii S., Tsuji A., Nagahama M., Imami A., Hayashi K.,		
RA Akamatsu T., Nagamune H., Matsuda Y.;		
RT "A novel human PACE4 isoform, PACE4E is an active processing protease		
RL containing a hydrophobic cluster at the carboxy terminus.";		
RL J. Biochem. 121:941-948(1997).		

RN [6]
 RP SEQUENCE FROM N.A. (ISOFORMS PACEA-I; A-II; CS; D; E-I; E-II).
 RX MEDLINE=99021085; PubMed=9378725;
 RA Tsuji A., Hine C., Tamai Y., Tomemoto K., Mori K., Yoshida S.,
 RA Bando M., Sakai E., Mori K., Akamatsu T., Matsuda Y.;
 RT "Genomic organization and alternative splicing of human PACE4 (SPC4),
 RT kexin-like processing endoprotease.";
 RL J. Biochem. 122:438-452(1997).
 RN [7]
 RP ALTERNATIVE SPLICING (ISOFORM PACE4CS).
 RX MEDLINE=97064242; PubMed=8906861;
 RA Zhong M., Benjannet S., Lazure C., Munzer S., Seidah N.G.;
 RT "Functional analysis of human PACE4-A and PACE4-C isoforms:
 RT identification of a new PACE4-CS isoform.";
 RL FEBS Lett. 396:31-36(1996).
 RN [8]
 RP CHARACTERIZATION.
 RX MEDLINE=99233559; PubMed=10215603;
 RA Susic J.F., Moehring J.M., Innocencio N.M., Luchini J.W.,
 RA Moehring T.J.;
 RT "Endoprotease PACE4 is Ca²⁺-dependent and temperature-sensitive and
 RT can partly rescue the phenotype of a furin-deficient cell strain.";
 RL Biochem. J. 339:639-647(1999).
 RN [9]
 RP PROCESSING.
 RX MEDLINE=98408849; PubMed=9738469;
 RA Nagahama M., Taniguchi T., Hashimoto E., Imamaki A., Mori K.,
 RA Tsuji A., Matsuda Y.;
 RT "Biosynthetic processing and quaternary interactions of proprotein
 RT convertase SPC4 (PACE4).";
 RL FEBS Lett. 444:155-159(1998).
 CC -1- FUNCTION: LIKELY TO REPRESENT AN ENDOPEPTIDASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY, WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF NATURE PROTEINS FROM THEIR
 CC PROTEOLYTIC BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-ZAA BONDS,
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- COFACTOR: PACE4 IS PROBABLY CALCIUM-DEPENDENT.
 CC -1- SUBUNIT: THE PACE4-I PRECURSOR PROTEIN SEEMS TO EXIST IN THE
 CC RETICULUM ENDOPLASMIC AS BOTH A MONOMER AND A DIMER-SIZED COMPLEX
 CC WHEREAS MATURE PACE4-I EXISTS ONLY AS A MONOMER, SUGGESTING THAT
 CC PROPEPTIDE CLEAVAGE AFFECTS ITS TERTIARY OR QUATERNARY STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: PACE4-I AND PACE4-AII ARE SECRETED. PACE4C
 CC AND PACE4S ARE NOT SECRETED AND REMAIN PROBABLY IN ZYMOGEN FORM
 CC IN ENDOPLASMIC RETICULUM. PACE4E-I AND PACE4E-II ARE RETAINED
 CC INTRACELLULARLY PROBABLY THROUGH A HYDROPHOBIC CLUSTER IN THEIR C-
 CC TERMINUS. PACE4B MIGHT BE SECRETED.
 CC -1- ALTERNATIVE PRODUCTS: 8 ISOFORMS: PACE4A-I/PACE4 (SHOWN HERE),
 CC PACE4A-II, PACE4B/PACE4.1, PACE4C, PACE4CS, PACE4D, PACE4E-I AND
 CC PACE4E-II; ARE PRODUCED BY ALTERNATIVE SPLICING. ISOFORMS PACE4B,
 CC C, CS AND D MIGHT BE ENZYMATICALLY INACTIVE.
 CC -1- TISSUE SPECIFICITY: EACH PACE4 ISOFORM EXHIBITS A UNIQUE
 CC RESTRICTED DISTRIBUTION. PACE4A-I IS EXPRESSED IN HEART, BRAIN,
 CC PLACENTA, LUNG, SKELETAL MUSCLE, KIDNEY, PANCREAS, BUT AT
 CC COMPARTMENTALLY HIGHER LEVELS IN THE LIVER. PACE4A-II IS AT LEAST
 CC EXPRESSED IN PLACENTA. PACE4B WAS ONLY FOUND IN THE EMBRYONIC
 CC KIDNEY CELL LINE FROM WHICH IT WAS ISOLATED. PACE4C AND PACE4D ARE
 CC EXPRESSED IN PLACENTA. PACE4E-I IS EXPRESSED IN CEREBELLUM,
 CC PLACENTA AND PITUITARY. PACE4E-II IS AT LEAST PRESENT IN
 CC CEREBELLUM.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZYMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM. ISOFORM PACE4D LACKS THE PROPEPTIDE DOMAIN.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC -1- SIMILARITY: CONTAINS 1 HOWO B/P DOMAIN.
 CC -----
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>)

CC or send an email to license@isb-sib.ch.
 CC -----
 DR EMBL; M80482; AAA5998.1; JOINED.
 DR EMBL; AB001914; BAA21620.1; JOINED.
 DR EMBL; AB001898; BAA21620.1; JOINED.
 DR EMBL; AB001900; BAA21620.1; JOINED.
 DR EMBL; AB001901; BAA21620.1; JOINED.
 DR EMBL; AB001902; BAA21620.1; JOINED.
 DR EMBL; AB001903; BAA21620.1; JOINED.
 DR EMBL; AB001904; BAA21620.1; JOINED.
 DR EMBL; AB001905; BAA21620.1; JOINED.
 DR EMBL; AB001906; BAA21620.1; JOINED.
 DR EMBL; AB001907; BAA21620.1; JOINED.
 DR EMBL; AB001908; BAA21620.1; JOINED.
 DR EMBL; AB001909; BAA21620.1; JOINED.
 DR EMBL; AB001910; BAA21620.1; JOINED.
 DR EMBL; AB001901; BAA21622.1; JOINED.
 DR EMBL; AB001902; BAA21622.1; JOINED.
 DR EMBL; AB001903; BAA21622.1; JOINED.
 DR EMBL; AB001904; BAA21622.1; JOINED.
 DR EMBL; AB001905; BAA21622.1; JOINED.
 DR EMBL; AB001906; BAA21622.1; JOINED.
 DR EMBL; AB001907; BAA21622.1; JOINED.
 DR EMBL; AB001908; BAA21622.1; JOINED.
 DR EMBL; AB001909; BAA21622.1; JOINED.
 DR EMBL; AB001910; BAA21622.1; JOINED.
 DR EMBL; AB001901; BAA21623.1; JOINED.
 DR EMBL; AB001902; BAA21623.1; JOINED.
 DR EMBL; AB001903; BAA21623.1; JOINED.
 DR EMBL; AB001904; BAA21623.1; JOINED.
 DR EMBL; AB001905; BAA21623.1; JOINED.
 DR EMBL; AB001906; BAA21623.1; JOINED.
 DR EMBL; AB001907; BAA21623.1; JOINED.
 DR EMBL; AB001908; BAA21623.1; JOINED.
 DR EMBL; AB001909; BAA21623.1; JOINED.
 DR EMBL; AB001910; BAA21623.1; JOINED.
 DR EMBL; AB001898; BAA21624.1; JOINED.
 DR EMBL; AB001900; BAA21624.1; JOINED.
 DR EMBL; AB001901; BAA21624.1; JOINED.
 DR EMBL; AB001902; BAA21624.1; JOINED.
 DR EMBL; AB001903; BAA21624.1; JOINED.
 DR EMBL; AB001904; BAA21624.1; JOINED.
 DR EMBL; AB001905; BAA21624.1; JOINED.
 DR EMBL; AB001906; BAA21624.1; JOINED.
 DR EMBL; AB001907; BAA21624.1; JOINED.
 DR EMBL; AB001908; BAA21624.1; JOINED.
 DR EMBL; AB001909; BAA21624.1; JOINED.
 DR EMBL; AB001910; BAA21624.1; JOINED.
 DR EMBL; AB001901; BAA21625.1; JOINED.
 DR EMBL; AB001902; BAA21625.1; JOINED.
 DR EMBL; AB001903; BAA21625.1; JOINED.
 DR EMBL; AB001904; BAA21625.1; JOINED.
 DR EMBL; AB001905; BAA21625.1; JOINED.
 DR EMBL; AB001906; BAA21625.1; JOINED.
 DR EMBL; AB001907; BAA21625.1; JOINED.
 DR EMBL; AB001908; BAA21625.1; JOINED.
 DR EMBL; AB001909; BAA21625.1; JOINED.
 DR EMBL; AB001910; BAA21625.1; JOINED.
 Query Match 11.45; Score 173.5; DB 1; Length 969;

FT CARBOHYD 804 804 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 854 854 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1710 1710 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 1732 1732 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT VARSPLIC 878 915 XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX ->
 FT VARSPLIC 878 915 AAEESMAEGFCMVLVKKNNLCCQKRYLOQLCCKCTFPOG
 FT VARSPLIC 878 915 (IN ISOFORM PCSA).
 FT VARSPLIC 878 915 MISSING (IN ISOFORM PCSA).
 SQ SEQUENCE 1877 AA; 20788 MW; 890955DC0534444 CRC64;

Query Match 11.1%; Score 168; DB 1; Length 1877;
 Best Local Similarity 26.3%; Pred. No. 0.0001;
 Matches 49; Conservative 25; Mismatches 54; Indels 58; Gaps 11;

QY 36 PNVSO-GCQG-----GCATCSYNGCLSCPKRLFALERIGKMGVGLSCSPGYGTRY 90
 DB 640 PECSEVCGDGPDPCHTCLHYHKLKNNTR-----ICVSSCPGPHF---H 682
 QY 91 PINKCTCKKADCDCTCFNKF--CTCKSGPYLH--LGKCLDNGPEGLANHTMECVSI 146
 DB 683 ADKRCRCACANCCSCGSHADCLCKYFLNETSSCVAOCPESSYODIKNIC--- 739
 QY 147 VHCSEVSEWNPSPCTKKGTC-GFKRGTESTRVRELIHPSAKGNLCPTNETRKTQVOR 205
 DB 740 -----GKCSNCKTCTGFHNCTE-----CKGGL---SLOGSRCSV--- 771
 QY 206 KCOGE 211
 DB 772 TCEDQ 777

RESULT 6
 PKCS_HUMAN
 ID PKCS_HUMAN STANDARD; PRT; 913 AA.
 AC 092824; Q13527;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Proprotein convertase subtilisin/kexin type 5 precursor (EC 3.4.21.-)
 DE (Protein convertase PCS) (Subtilisin/kexin-like protease PCS)
 DE (Convertase PCS) (PC6) (hPC6).
 GN PCSK5 OR PCS OR PC6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=T-cell;
 RA MEDLINE=96353880; PubMed=8755538;
 RA Miranda L., Wolf J., Pichuanes S., Duke R., Franzusoff A.;
 RT "Isolation of the human PC6 gene encoding the putative host protease
 for HIV-1 gp160 processing in CD4+ T lymphocytes.";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:7695-7700(1996).
 RN [2]
 RP REVISIONS
 RA Franzusoff A., Miranda L., Wolf J., Pichuanes S., Lu Y., Duke R.;
 RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE OF 15-913 FROM N.A.
 RA Redelhuber T.L.;
 CC Submitted (FEB-1996) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: LIKELY TO REPRESENT A WIDESPREAD ENDOPEPTIDASE ACTIVITY
 CC WITHIN THE CONSTITUTIVE AND REGULATED SECRETORY PATHWAY. CAPABLE
 CC OF CLEAVAGE AT THE RX(R/R)X CONSENSUS MOTIF.
 CC -1- CATALYTIC ACTIVITY: RELEASE OF MAJORE PROTEINS FROM THEIR
 CC PROPOTERINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-1-ZAA BONDS, WHERE XAA
 CC CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC -1- SUBCELLULAR LOCATION: Secreted (by similarity).
 CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS SEEMS TO BE PRODUCED BY
 CC ALTERNATIVE SPLICING.
 CC -1- TISSUE SPECIFICITY: EXPRESSED IN T-LYMPHOCYTES.
 CC -1- DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE

CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC -1- SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; U56387; AAC50643.2; -
 CC EMBL; U49114; AAA91807.1; -
 CC HSSP; O99405; 1MPT.
 CC MEROPS; S08.076; -
 CC Genew; HGNC:8747; PCSK5.
 CC MIM; 600488; -
 DR InterPro; IPR002174; Furin-like.
 DR InterPro; IPR002884; P_domain.
 DR InterPro; IPR002029; Peptidase_S8.
 DR Pfam; PF00082; Peptidase_S8; 1.
 DR Pfam; PF01483; P_PARTIAL.
 DR PRINTS; PR00723; SUBTILISIN.
 DR Prodom; PD000717; P_domain; 1.
 DR SMART; SM00261; FU_5.
 DR PROSITE; PS00136; SUBTILASE ASP. 1.
 DR PROSITE; PS00137; SUBTILASE_HIS. 1.
 DR PROSITE; PS00138; SUBTILASE_SER. 1.
 DR Hydrolase; Serine protease; Glycoprotein; Zymogen; Signal;
 KW Cleavage on pair of basic residues; Repeat.
 FT SIGNAL 1 32
 FT PROPEP 33 114
 FT CHAIN 115 913
 FT TYPE 5.
 FT DOMAIN 115 454
 FT DOMAIN 454 600
 FT DOMAIN 600 913
 FT SITE 114 115
 FT SITE 519 521
 FT ACT SITE 171 171
 FT ACT SITE 212 212
 FT ACT SITE 386 386
 FT CARBOHYD 225 225
 FT CARBOHYD 381 381
 FT CARBOHYD 665 665
 FT CARBOHYD 752 752
 FT CARBOHYD 802 802
 FT CARBOHYD 852 852
 FT CONFLICT 118 118
 FT CONFLICT 121 121
 FT CONFLICT 511 511
 FT CONFLICT 601 601
 SQ SEQUENCE 913 AA; 101775 MW; 21389264C4AD7546C CRC64;

Query Match 10.7%; Score 162; DB 1; Length 913;
 Best Local Similarity 22.3%; Pred. No. 0.0002;
 Matches 55; Conservative 33; Mismatches 81; Indels 78; Gaps 12;

QY 21 GSNASRGROPRMHPVSOCCGATCSYNG--CLSCPKRLFALERIGKMGVGL 78
 DB 677 GHYHDK-KRCCKAPN-----CESCFGSHGDCMCKYGFLL-----NEINSCV 721
 QY 79 SSCPSC-YGGRYPDINKCTKCAADDCTCFNKFCTCKSGPYLHKLKCLDNGPEGLAN 137
 DB 722 THCPDSYODTKK--NLCKRCSENCCTCFEHNCTEGRDLISLOGSRCSVSEDRYEN 778
 QY 138 NHTME-----CVS-----YHCVSEW-----NFWSPCTCK 163
 DB 779 GDCPCRPATCAGAGAGCINCTEGRFMEGRCVQSGSISVYFPHSENGYKCKKC 838
 QY 164 GTC-----GFKRGTE-----TRVRELIHPSAKGNLCPTNETRKTCTV 202

Db 839 DISCTCNGGFGFKCTSCPSGYLLDGMCGMCAICXKATATESNAGEGFCMLVKXNNLC-- 896
 Oy 203 ORKCKQK 209
 Db 897 ORKVLQK 903

RESULT 7
 WIF1_HUMAN STANDARD; PRT; 379 AA.
 AC 09Y5M5;
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Wnt inhibitory factor 1 precursor (WIF-1).
 GN WIF1.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE=99215557; PubMed=10201374;
 RA Hsieh J.-C., Kodjabachian L., Rebber M.L., Rattner A.,
 RA Smallwood P.M., Samos C.H., Nusse R., David I.B., Nathans J.;
 RT "A new secreted protein that binds to Wnt proteins and inhibits their
 RT activities.";
 RT Nature 398:431-436(1999).
 RL Nature 398:431-436(1999).
 CC - FUNCTION: BINDS TO WNT PROTEINS AND INHIBITS THEIR ACTIVITIES. MAY
 CC BE INVOLVED IN MESODERM SEGMENTATION.
 CC - SUBCELLULAR LOCATION: Secreted.
 CC - SIMILARITY: CONTAINS 5 EGF-LIKE DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 CC EMBL; AF129222; AAD25402.1; -;
 DR HSSP; P00743; IICF.
 DR Genew; HGNC:18081; WIF1.
 DR MIM; 605186; -;
 DR InterPro: IPR000561; EGF-like.
 DR InterPro: IPR002049; Laminin_EGF.
 DR InterPro: IPR003306; WIF.
 DR Pfam; PF00008; EGF; 5.
 DR Pfam; PF02019; WIF; 1.
 DR PRINTS; PR00011; EGF_LAMININ.
 DR SMART; SM00181; EGF; 5.
 DR SMART; SM00469; WIF; 1.
 DR PROSITE; PS00022; EGF_1; 5.
 DR PROSITE; PS01186; EGF_2; 4.
 KM Repeat: EGF-like domain; Signal; Developmental protein.
 FT CHAIN 1 28
 FT DOMAIN 1 177 208 WNT INHIBITORY FACTOR 1.
 FT DOMAIN 2 209 240 EGF-LIKE 1.
 FT DOMAIN 3 241 272 EGF-LIKE 2.
 FT DOMAIN 4 273 304 EGF-LIKE 3.
 FT DOMAIN 5 305 336 EGF-LIKE 4.
 FT DISULFID 177 186 EGF-LIKE 5.
 FT DISULFID 182 192 POTENTIAL.
 FT DISULFID 198 200 POTENTIAL.
 FT DISULFID 209 218 POTENTIAL.
 FT DISULFID 214 224 POTENTIAL.
 FT DISULFID 230 232 POTENTIAL.
 FT DISULFID 241 250 POTENTIAL.
 FT DISULFID 246 256 POTENTIAL.
 FT DISULFID 262 264 POTENTIAL.

FT DISULFID 273 282 POTENTIAL.
 FT DISULFID 278 288 POTENTIAL.
 FT DISULFID 294 296 POTENTIAL.
 FT DISULFID 305 314 POTENTIAL.
 FT DISULFID 310 320 POTENTIAL.
 FT DISULFID 326 328 POTENTIAL.
 FT CARBOHYD 88 88 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 FT SEQUENCE 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
 SQ SEQUENCE 379 AA; 41512 MW; 27782370A266E784 CRC64;

Query Match 10.54; Score 158.5; DB 1; Length 379;
 Best Local Similarity 24.94; Pred. No. 0.00017;
 Matches 60; Conservative 26; Mismatches 72; Indels 83; Gaps 16;

Oy 42 CQGGCT--CSDVNGC-----LSKRLPFLERIGMKQIGYCSS-----C 81
 Db 182 CPGGNGGFCNERRICCPDPFGPHCEKALCTPCNNG-----GLCVPGFCIC 232
 Oy 82 PSQYVTRPYDINKTKC-KADCD-TCFNKFCFKCKSGFYLLHKLCDNCEGLNANN 139
 Db 233 PPGFYG-----VNCCKXKNCSTTCFNGTC-----FY--PKCI--CPGLEGE-- 271
 Oy 140 TMECVIVHCEVSEWNPSPCTKKGTCGFGKGTETRYREIIQHSKAGNLCPP----- 193
 Db 272 -----OCEISKCP--QPCNNGKCTG---KSKCKSGKYQGDLCSPVCEPGCGANG 318
 Oy 194 -TWETKCTVQRKCKCKGCKGKGRERK-----KPNKESKKAIPDSKLSSEKPEIP 246
 Db 319 TCEPRKCK-----CQEGMGHGRHNKRYEALSHLRPAGQLRQHTPSLKAKKEERRDP 373
 Oy 247 E 247
 Db 374 E 374

RESULT 8
 PAC4_RAT STANDARD; PRT; 937 AA.
 AC 063415;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Paired basic amino acid cleaving enzyme 4 precursor (EC 3.4.21.-)
 DE (Subtilisin/kexin-like protease PACE4) (Subtilisin-like proprotein
 DE convertase 4) (SPC4).
 GN PACE4.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_TaxID=10116;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=Sprague-Dawley; TISSUE=Hypothalamus, and Pituitary;
 RX MEDLINE=94349873; PubMed=8070361;
 RA Johnson R.C., Darlington D.N., Hand T.A., Bloomquist B.T., Mains R.E.;
 RT "PACE4, a subtilisin-like endoprotease prevalent in the anterior
 RT pituitary and regulated by thyroid status.";
 RT Endocrinology 135:1178-1185(1994).
 RL Endocrinology 135:1178-1185(1994).
 CC - FUNCTION: LIKELY TO REPRESENT AN ENDOPEPTASE ACTIVITY WITHIN THE
 CC CONSTITUTIVE SECRETORY PATHWAY. WITH UNIQUE RESTRICTED
 CC DISTRIBUTION IN BOTH NEUROENDOCRINE AND NON-NEUROENDOCRINE TISSUES
 CC AND CAPABLE OF CLEAVAGE AT THE RX(K/R)R CONSENSUS MOTIF.
 CC - CATALYTIC ACTIVITY: RELEASE OF MATURE PROTEINS FROM THEIR
 CC PROPORTEINS BY CLEAVAGE OF ARG-XAA-YAA-ARG-|-2AA BONDS.
 CC WHERE XAA CAN BE ANY AMINO ACID AND YAA IS ARG OR LYS.
 CC - COFACTOR: PROBABLY CALCIUM-DEPENDENT (BY SIMILARITY).
 CC - TISSUE SPECIFICITY: HIGH EXPRESSION IN THE ANTERIOR PITUITARY AND
 CC IN SEVERAL BRAIN REGIONS, THE ATRIUM, AND THE VENTRICLE.
 CC - DOMAIN: THE PROPEPTIDE DOMAIN ACTS AS AN INTRAMOLECULAR CHAPERONE
 CC ASSISTING THE FOLDING OF THE ZMOGEN WITHIN THE ENDOPLASMIC
 CC RETICULUM.
 CC - SIMILARITY: BELONGS TO PEPTIDASE FAMILY S8.
 CC - SIMILARITY: CONTAINS 1 HOMO B/P DOMAIN.